

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 5, 2003, 17:54:18 ; Search time 1646.7 Seconds
(without alignments)
5302.013 Million cell updates/sec

Title: US-09-990-099-21

Perfect score: 300
Sequence: 1 attcatcgtgtcgtcgtcgtg.....gattccagagactatgatt 300

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapept 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

GenBml:*

- 1: gb_da:*
- 2: gb_hlg:*
- 3: gb_in:*
- 4: gb_om:*
- 5: gb_ov:*
- 6: gb_pat:*
- 7: gb_ph:*
- 8: gb_pl:*
- 9: gb_pt:*
- 10: gb_ro:*
- 11: gb_sts:*
- 12: gb_sy:*
- 13: gb_un:*
- 14: gb_vi:*
- 15: em_ba:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_mu:*
- 20: em_om:*
- 21: em_or:*
- 22: em_ov:*
- 23: em_pat:*
- 24: em_ph:*
- 25: em_pl:*
- 26: em_ro:*
- 27: em_sts:*
- 28: em_un:*
- 29: em_vi:*
- 30: em_hlg_hum:*
- 31: em_hlg_inv:*
- 32: em_hlg_other:*
- 33: em_hlg_mus:*
- 34: em_hlg_pln:*
- 35: em_hlg_rod:*
- 36: em_hlg_mam:*
- 37: em_hlg_vrt:*
- 38: em_sy:*
- 39: em_hgo_hum:*
- 40: em_hgo_mus:*
- 41: em_hgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	300	100.0	720	1	AF407018	AF407018 Escherich
C 2	300	100.0	2886	1	ECOPR0TS	M94104 Escherichia
C 3	300	100.0	10120	1	AE000446	AE000446 Escherich
C 4	300	100.0	136254	1	ECODW82	L100328 E. coli: th
C 5	300	100.0	252977	2	AC073670	AC073670 Mus muscu
C 6	300	100.0	256377	2	AC020870	AC020870 Mus muscu
C 7	300	100.0	272545	2	AC090533	AC090533 Mus muscu
C 8	298.4	99.5	12542	1	AE005600	AE005600 Escherich
C 9	298.4	99.5	280900	1	AP002566	AP002566 Escherich
C 10	295	98.3	221285	2	AC023611	AC023611 Mus muscu
C 11	184.4	61.5	245050	1	AL627280	AL627280 Salmoneil
C 12	182.8	60.9	22492	1	AE008877	AE008877 Mus muscu
C 13	36.8	12.3	181977	2	AC103606	AC103606 Oryza sat
C 14	36.6	12.2	160174	2	AP005103	AP005103 Oryza sat
C 15	36.2	12.1	195007	9	CNS06C81	AL391749 Human chr
C 16	36.2	12.0	108822	9	AC004968	AC004968 Homo sapi
C 17	36	12.0	178201	2	AC019189	AC019189 Homo sapi
C 18	36	12.0	207942	9	AC069499	AC069499 Homo sapi
C 19	35	11.7	139534	2	AP005640	AP005640 Oryza sat
C 20	34.6	11.5	1380	10	AF232010	AF232010 Cavia sp.
C 21	34.6	11.5	137426	2	AC079357	AC079357 Oryza sat
C 22	34.6	11.5	257214	2	AC126248	AC126248 Mus muscu
C 23	34.6	11.5	299480	9	AC121563	AC121563 Mus muscu
C 24	34.4	11.5	1853	8	AK000917	AK000917 Homo sapi
C 25	34.4	11.5	85413	8	AC009324	AC009324 Arabidops
C 26	34.4	11.5	153772	2	AC104464	AC104464 Homo sapi
C 27	34.4	11.5	158898	2	AL359879	AL359879 Homo sapi
C 28	34.4	11.5	171482	9	AC034247	AC034247 Homo sapi
C 29	34.4	11.5	171961	9	AC113408	AC113408 Homo sapi
C 30	34.4	11.5	200175	2	AC096004	AC096004 Rattus no
C 31	34	11.3	78376	9	AL359692	AL359692 Human DNA
C 32	34	11.3	95836	9	AC016710	AC016710 Homo sapi
C 33	33.8	11.3	128597	2	AC098295	AC098295 Rattus no
C 34	33.8	11.3	164879	2	AC007876	AC007876 Homo sapi
C 35	33.8	11.3	186747	9	AC018605	AC018605 Homo sapi
C 36	33.8	11.3	199913	9	AC007495	AC007495 Homo sapi
C 37	33.6	11.2	122379	9	AL359715	AL359715 Human DNA
C 38	33.6	11.2	151040	2	AC068389	AC068389 Homo sapi
C 39	33.6	11.2	161595	2	AC012267	AC012267 Homo sapi
C 40	33.6	11.2	175263	2	AC023102	AC023102 Homo sapi
C 41	33.6	11.2	176856	2	AC124145	AC124145 Rattus no
C 42	33.4	11.1	98494	9	AC087763	AC087763 Homo sapi
C 43	33.4	11.1	125020	9	AF429315	AF429315 Homo sapi
C 44	33.4	11.1	176684	2	AC083927	AC083927 Homo sapi
C 45	33.4	11.1	185510	2	AC011018	AC011018 Homo sapi

ALIGNMENTS

RESULT 1
AF407018/c
LOCUS AF407018
DEFINITION Escherichia coli inner membrane protein (ecfI) gene, complete cds.
ACCESSION AF407018
VERSION AF407018.1 GI:15529642
KEYWORDS
SOURCE
ORGANISM Escherichia coli.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia coli.
REFERENCE
AUTHORS 1 (bases 1 to 720)
TITLE Darligalouque, C., Missiakas, D. and Raina, S.
JOURNAL Characterization of the Escherichia coli sigma E regulon
J. Biol. Chem. 276 (24), 20866-20875 (2001)

MEDLINE 21293005
 PUBMED 11274153
 REFERENCE 2 (bases 1 to 720)
 AUTHORS Dartigalongue, C., Msiaklas, D. and Raina, S.
 TITLE Direct Submission
 JOURNAL Submitted (09-AUG-2001) Blochimie Medicale, CMT, 1 Rue Michel-

FEATURES	Location/Qualifiers
source	1. .720

```

gene      /organism="Escherichia coli"
          /db_xref="taxon:562"
          142. >694
promoter  /gene="ecfI"
          142..177

```

```

/ gene="ecfI"
/ note="factor sigma24"
178. .>694
/ gene="ecfI"
/ product="inner membrane protein"
287. .694
CDS

```

```

/genes="ecf1"
/notes="Ecfl; under heat shock transcriptional control"
RpoE"

```

BASE COUNT	207	a	169	c	183	g	161	t
ORIGIN	FRKSSSVRSREKSEANAQATNAVIPARMPDN"							

Query Match	Score	DB	Length
Best Local Similarity	100.08	1	720
Matches 300; Conservative	0	4,1e-87	0
Mismatches	0	0	0

Df

07
1 ATTCATCTGTGATCGTGGGTGTGGCCGAGTAGACTATTACGATCCTTGCTCAATAAATA 60
|||
356 ATTCAATCTGTGATCGTGGGTGTGGCCGAGTAGACTTTAGAGAAACCCTCCCCAAAATA 307

61 ACATCATCATACGTGCGACTTGGCGGCTATCGACATTTAAGTTTGTGCTGCCCCCT 120

09 121 CAGTCTATGCATAGACCATTAACCTGCAAAAAAGTCGCTGATTAAGGCTTGAAAAAGT 180

06 296 ACATCATTTCATTCAGTGCACACTGTGGGCGCTATGCGACACTTTAAAGCTTTGTCGTGCCCCCT 237

[illegible]

101	CATTCCAGACCCATTTTACATCGTAGCCGATGAGGACGGCCCTGATGGGTCTTCGGC	240
176	CATTTCAGACCCATTTTACATCGTAGCCGATGAGGACGGCCCTGATGGGTCTTCGGC	117

QY 241. TACCGACCCTGCATTTGTGGAAGCTTACATTTCGCTGATTTTCAGAGCAATTGATTT 300
|||||
Db 116 TACCTGACCTCCTCCATTCCTGCTGATTTTCAGAGCAATTGATTT .

[illegible]

ECOPROTS	LOCUS	DEFINITION	2886 bp	DNA	linear	BCT 26-APR-1993
ECOPROTS	LOCUS	Escherichia coli ibpa and ibpb genes, complete cds.				
		M91104				

VERSION	M94104.1	GI:147368	.	.
KEYWORDS	ibpA gene; ibpB gene.			
SOURCE	Escherichia coli (sub strain W3110	strain W-13)	nm	

ORGANISM
Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.

REFERENCE	AUTHORS	TITLE
1 (bases 1 to 2886)	Allen, S.P., Polazzi, J.O., Gierse, J.K. and Easton, A.M.	Two novel heat shock genes encoding proteins produced in response to heat shock

JOURNAL
of heterologous protein expression in *Escherichia coli*
t. Bacteriol. 174 (21), 6938-6947 (1992)
MEDLINE
93015757
PUBMED
1386969

source 1. .2886

```

/organism="Escherichia coli"
/strain="K-12"
/sub_strain="W3110"
/db_xref="taxon:562"
539..1063
gene
-gene="lppA"
539..545
-35_signal

```

```
-10_signal
/gene="ibpa"
/note="putative"
560..566
/gene="ibpa"
```

```

RBS      /note="putative"
          656..660
          /gene="lbpa"
          /note="putative"
CDS       670..1083

```

gene
1181. 1623

```

RBS
1181. 1186
/gene="1bpb"
/note="putative"

```

```

CDS
    1195..1623
        /gene="lbpB"
        /note="putative"
        /codon_start=1

```

```

/transl_table=11
/protein_id="AAA24425.1"
/db_xref="GI:147370"

```

```

/translation="MRNPDISPLMKQWIGCDKLANALQNGESQSPPYNIKESDDNH
YRTLLALAGFRQEDLEIQLEGTRSVKGTPEQKEKRTKTHQGLMQPFSLSFTLAE
MEVSGATFENGLIHIDLIRNEPEPIAARATISEPAINC"

```

terminator	1639. .1666
stem_loop	/note="putative" 1639. .1659

```

-35_signal      /note="putative"
                1769; .1774
                /note="putative"

```

-10_signal	1791.1796	/note="putative"
RAS	1805.1809	

```

CDS
    /note="putative"
    1819..2862
    /note="ORF; putative"
    /codon_start=1

```

```
/codon_scale=1  
/transl_table=11  
/protein_id="AAA24426.1"  
/db_xref="GI:147371"  
/translation="MGRKLRNKKR"
```

7/CLANSIALION=MSDALNVLALAVWGFIGNKFRGICIGIGVLFEGITVG
HFVSQAGMTLSDDMLHVIOEFGILFVYTTIGIQVPGPFASLRVGLRLNLFVLI
IGCLVAILHKLFEDIPVLVGLFSGAVTNPALGAGQQLRLDGTPEHNVDDGMST
AMAYPGICIGILETMMRPLVRDNRNRRALRRGALRRGALRRGALRRGALRRG

ISIORDRNATI."#
EVDSTSTKTDLIVERVVVNENUGKRIRDLHFKERYDVVYISRLNFAVGLVAVSG
LAIKDVPILNGDKIICSRLEKREDTLKVSPDTIQLGDLHLHGOPADLHNAOLIGQ
IICQZELIMMUNVAIRVNELEAQHSSSRINGALIKITINRVENPULHD

BASE COUNT	693 a	701 c	722 g	770 t
ORIGIN				

```
Query Match      100.0%; Score 300; DB 1; Length 2886;
```


/note="f416; 100 pct identical amino acid sequence and equal length to YDR_ECOLI SW: P31455"

/codon_start=1

/transl_table=11

/product="orf, hypothetical protein"

/protein_id="AAC76712.1"

/db_xref="GI:1790124"

/translation="MAGPVLVYQDRAMKQITFAPRNHLLTNTMTTPSPQWLVDPVRSQSFGEITIEVNIHTGEVEYIRASQAGAVGVTVKSEKVFTHGHPNDETHHEDPALDRNVGVAAPGCVNVOKOHREISGSHCVLVSRTPTDQSGDELNAVREGVGNHAIAPIGITLSPKGRPELTVLPODEAKVAKAGAPLSGTELLPAPRGVYORITFTTHRAIPGLVAVPRHWACNPGQIOIAFLMRDNDGIVOLMLISQSGEPROLHNKTIDIOSAENMHPSEGMLGFVLDNRACAHQSGEVEYTFEHANSPSADAVVSPDQMLAMWEGQIMITEEDR"

2359..2388

/note="factor sigma70; predicted +1 start at 3866948"

2430..3515

/gene="yids"

/note="b3690"

2430..3515

/gene="yids"

/function="orf; Unknown"

/note="o361; 100 pct identical amino acid sequence and equal length to YIDS_ECOLI SW: P31456"

/codon_start=1

/transl_table=11

/product="orf, hypothetical protein"

/protein_id="AAC76713.1"

/db_xref="GI:1790125"

/translation="MMNVSEKMEHFDVAIIGPAGSALARKAKQVIALKKHOCGTEGFSKPCGGLAPDAORSEITGDLTPDVLAVNPQIFSVKTVDAASLITRYQSYINIRHAFDLMMKSLIPASVEYHDSICRTKEDDWHVIFRADVGEQHTIARYLVGPIRGATAPKDGQRTFTLKEKSAFOQFGKTVKSEKTVLPSRMODEVCCKDNVLIIEAGFISASSLEGISYALDSIDIRSVLLAQPEKLTAVYRATRKRLKLFKIVKSRCLTAPALRKIMRSGVAHIFOLKD"

complement(2436..2464)

/note="factor sigma70; predicted +1 start at 3866982"

complement(2524..2552)

/note="factor sigma70; predicted +1 start at 3867070"

complement(3512..4849)

/gene="dgot"

/note="b3691"

complement(3512..4849)

/gene="dgot"

/function="transport; Transport of small molecules: Carbohydrates, organic acids, alcohols"

/note="f445; formerly designated yidU"

/codon_start=1

/transl_table=11

/product="D-galactonate transport"

/protein_id="AAC76714.1"

/db_xref="GI:1790126"

/translation="WYSGFRMPKIRKRLAMDIPVNAKGRHRYLLVWIFTVICYVBRANLAVASAHIOEFGITKAMGVESAFAMLYLCOIPGMFLDRGSRVAFITGWSVALTEFGFATGLMSLIGIRITGFEAPAPNTNRMTSMPEPHERSAVFIATGQFVGLAFLLPLIMIOEMLSMHWIVTGIGITLIMLEKRYOQPRRLTKGSKAELDYIDGGGLVDGAPVKEAROPLTAKMDKLVHRKILGYLQDRAVASTLWFLNPPNYLQEKGTALAKAGMTVPPLAFLVGLSGVADLVLRKGFSLGFAKRTPIICGLISTCIMGANVTNDPMIMCMLAALFQNGFASITWSSVSLAPMLIGLGGVNFAGLIGITVPLVYGYLAQGGFAPALVYISAVALLGALSYLLVGVKRVG"

complement(4924..6687)

/gene="dgoa"

/note="b3692"

complement(4924..6687)

/gene="dgoa"

/function="enzyme; Degradation of small molecules: Carbon compounds"

/note="f587; formerly designated yidU"

/codon_start=1

/transl_table=11

/product="2-oxo-3-deoxygalactonate 6-phosphate aldolase"

Query Match 100.0%; Score 300; DB 1; Length 10120;
Best Local Similarity 100.0%; Pred No. 5.3e-87;
Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

and galactonate dehydratase"
/protein_id="AAC76715.1"
/db_xref="GI:2367262"

/translation="MOMQRLPLILRLITGTPDEALAHGAVIDAFDAVEIPLNSPQMEOSIRPAVDYDKALIGAGTVLKEPQVDALARMCCOLIVPINHSEVIRAVGYMTCPGCATATAEFTALEAGAAAGENISVIGFMSAIHQSVKSGIAIGHRLICMRDAENLQAWIDAGCAGAGLSLDYIRAOQSVERTAQQAARFVKAYREAGMALKIKITTYRLP

QY 1 ATTCATCTGTTGATCGTGGGTTGGCTGATGATTAGCATCCCTGCTGAANA 60
DB 796 ATTCATCTGTTGATCGTGGGTTGGCTGATGATTAGCATCCCTGCTGAANA 737
QY 61 ACATCATCATTCAGTCCGACGTGGGCGATGCGACTTTAGCTTTCGTCGCCCT 120
DB 736 ACATCATCATTCAGTCCGACGTGGGCGATGCGACTTTAGCTTTCGTCGCCCT 677
QY 121 CAGCTATGCATAGACCATTAACCTGCAAAAAAGTCCGCTGATAAGCTTGAAGA 180
DB 676 CAGCTATGCATAGACCATTAACCTGCAAAAAAGTCCGCTGATAAGCTTGAAGA 617
QY 181 CATTCAGACCATTTTACATCGTAGCCGATGAGAGCGGCTGATGGCTTGTG 240
DB 616 CATTCAGACCATTTTACATCGTAGCCGATGAGAGCGGCTGATGGCTTGTG 557
QY 241 TACCTGACCTGTCATTTGAGAGGCTTACATTCGCTGATTCAGAGCATTTGAT 300
DB 556 TACCTGACCTGTCATTTGAGAGGCTTACATTCGCTGATTCAGAGCATTTGAT 497

RESULT 4
LOCUS ECOLW82/c 136254 bp DNA linear ECT 07-FEB-1995
DEFINITION E. coli; the region from 81.5 to 84.5 minutes.
ACCESSION L10328
VERSION L10328.1 GI:290484
KEYWORDS
SOURCE Escherichia coli K12 strain MG1655; lambda clones EC14-52, EC17-187, EC27-890, EC21-99, EC22-175, EC27-236, EC17-30, EC17-137, EC19-71, EC15-15, EC17-159; subclones in M13mp19 or Janus.
ORGANISM Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.

REFERENCE 1 (bases 1 to 136254)
AUTHORS Burland, V., Plunkett, G. III, Daniels, D.L. and Blattner, F.R.
TITLE DNA sequence and analysis of 136 kilobases of the Escherichia coli genome: organizational symmetry around the origin of replication
JOURNAL Genomics 16 (3), 551-561 (1993)
MEDLINE 93315143
PUBMED 7686882

COMMENT This sequence was determined as part of the E. coli Genome Project (Frederick R. Blattner, director) at the University of Wisconsin-Madison. Supported by award HG00301 from the NIH Human Genome Project. A preliminary report was presented at the Hilton Head meeting 'Genome Sequencing and Analysis IV' September 26-30, 1992. The entire sequence was independently determined from E. coli MG1655; overlaps and conflicts with other sequence determinations are annotated. The end of this entry overlaps the start-of the entry ECOLW85 (M87049) by the six bases of an EcoRI site.

Data kindly submitted in computer readable form by: Guy Plunkett III

Laboratory of Genetics
University of Wisconsin
445 Henry Mall
Madison WI 53706
USA
Phone: 608-262-2534
Email: ecoligenetics.wisc.edu
Fax: 608-263-7459

FEATURES [1] Authors request hold until publication.

Location/Qualifiers
 source 1..136254
 /organism="Escherichia coli"
 /db_xref="taxon:562"
 misc_feature 1..11607
 /note="EC14-52 (lambda clone)"
 /complement(<1..669)
 /note="X06036; ECORPG(1..>669)"
 /complement(1..544)
 /gene="fpg"
 /complement(<1..544)
 /gene="fpg"
 CDS
 /gene="alternate gene name mutH"
 /note="start=1"
 /codon_start=1
 /transl_table=11
 /product="Formamidopyrimidine-DNA glycosylase"
 /protein_id="AA61988.1"
 /db_xref="GI:643609"
 /translation="MPPLPEVETSRNGIEPHLVGATILHAHVENGRLMPVSEIYRLSDPVLVSORAKYLLLEPEGMWIIHGMGSRILPELPPEKHDHVDLVNSGKYLRTDPPRGAMWTKELGHNVLTHGPEPLSDDFNGEYLHOKCAKKTAKIPMLMDNKLVGNGNIYASSTLFACT"
 /complement(526..1290)
 /note="J01677; ECORPMB(1..764)"
 /complement(557..586)
 /note="promoter-like sequence; promoter matrix score of 44"
 misc_difference 596..601
 /note="AAAAA in X06036; AAAAAA in J01677 and here"
 misc_difference 611
 /note="G in J01677; A in X06036 and here"
 misc_difference 615
 /note="G in J01677; A in X06036 and here"
 /note="TTTTTTT in X06036; TTTT in J01677 and here"
 /complement(642..809)
 /gene="rpmg (CG Site No. 236)"
 /complement(642..809)
 /gene="rpmg (CG Site No. 236)"
 /codon_start=1
 /transl_table=11
 /product="50S ribosomal subunit protein L33"
 /protein_id="AA61989.1"
 /db_xref="GI:290486"
 /translation="MAKGIREKIKVSSAGTGHPYTTTKNKKRKEKLELKDFEVR OHVYREAKIK"
 /complement(830..1066)
 /gene="rpmB (CG Site No. 240)"
 /complement(830..1066)
 /gene="rpmB (CG Site No. 240)"
 /codon_start=1
 /transl_table=11
 /product="50S ribosomal subunit protein L28"
 /protein_id="AA61990.1"
 /db_xref="GI:290487"
 /translation="MSRVCOVTGKRPVTCGNNRSHALNATKRRLPLNLSHRPVESEK RPYTLKVSAGMVIDKGIIDVLAELRARGEKY"
 misc_difference 902
 /note="G in J01677; A here"
 misc_difference 1128
 /note="TC in J01677; C here"
 /note="T in J01677; TAT here"
 /complement(1168..1819)
 /note="X63366; ECORADC(1..651)"
 /complement(1212..1240)
 /note="promoter-like sequence; promoter matrix score of 77"
 gene
 /complement(1283..1957)
 /gene="radC (CG Site No. 13913)"
 /complement(1283..1957)
 /gene="radC (CG Site No. 13913)"
 CDS

/note="differs from radC sequence in X63366; similar to S. aureus putative 25 kD protein"
 /codon_start=1
 /transl_table=11
 /protein_id="AA61991.1"
 /db_xref="GI:290488"
 /translation="MKVKNNSOLMPREKMLKRGISALNDVELLALFRTGRGKDVLTAKEMLENGSLYGLTSEYEPGSGVHGIGAKFMOLKGLAEVLRVYVNMRESEPLASEPTEFLQSOLTEEREITMVLFLDSORHVTYTRHFLSGTLNHYVHREIYREAIKINASALILAHNHPSGCAEPSKADKLITERIKSCQMDLRVDHYIGGEYVSFAERGMI"
 /note="A in X63366; AA here"
 misc_difference 1369..1370
 /note="A in X63366; AA here"
 variation 1436
 /note="A deleted in radC102 mutation (X63367)"
 1698..1822
 /note="predicted bend of 76 degrees"
 1875..1902
 /note="promoter-like sequence; promoter matrix score of 50"
 /complement(1911..1940)
 /gene="radC (CG Site No. 13913)"
 /note="promoter-like sequence; promoter matrix score of 45"
 2051..3343
 /gene="dtp (CG Site No. 18430)"
 2051..3343
 /gene="dtp (CG Site No. 18430)"
 /codon_start=1
 /transl_table=11
 /protein_id="AA61992.1"
 /db_xref="GI:290489"
 /translation="MKARQKYCKDIANFCHPTGKTIIMSLAGKVIYVSGGTAAYKPELVRLRLDRGADVAVAMENAKAFITPLSLQAVSGIPVSDLLDRAEAKMHTLNLVADVLAPATADILARVAGAMANDVSTCLATPAPVAVLPMNOQMTAAATQHNIEVLASRGLIIMGPSGSOACGDIQPRAXDPLITVMAVAHFSVPVNDLKHLMITACGTEPRDLPVYISNHSKGMGFALIAAARAGANVYLVGVSPLTPPPKRVDAVTALEMAVYMAVSOONIFITGCAAVADRAATVAPEKIKKOAGDELTITMKVNPDI VAGVAALKDHRPYVVGFAEPYNNVEYVROKRIKRNLDILANDVSCOTQCFNSDNNALHLFWQDGRVPLELKEKELGQILDDELVTRIDENRR"
 /note="V01578; ECODUTPYR(1..2568)"
 2982..4591
 /note="X01714; ECOPYREA(1..1609)"
 3324..3779
 /gene="dut (CG Site No. 832)"
 3324..3779
 /gene="dut (CG Site No. 832)"
 /codon_start=1
 /transl_table=11
 /product="deoxyuridine triphosphatase"
 /protein_id="AA61993.1"
 /db_xref="GI:290490"
 /translation="MKKIDVKIILDPYKGFPLPYATSGSAGLDRACLANDAVELAPGDTLVPTGLAIHIDAPSLAAMLPISGLCHKRGIVLGNVGLIDSDYGOGLMISWNGQDSFTIOPGERIAQIMFVYVQAEINLVEDDADVBGGGHSGRQ"
 3844..4482
 /gene="ttk"
 3844..4482
 /gene="ttk"
 /function="unknown"
 /codon_start=1
 /transl_table=11
 /product="23 kD protein"
 /protein_id="AA61994.1"
 /db_xref="GI:290491"
 /translation="MPGKCLFSGVPCNMAEKONAKNRREILISALMLSESSDGO RTTAKLAASGVSEALYRHPSKTRPDSLIEFTDSITRTINILINKDKXTARLRLIYLILGFCERNGLRILITGHALMFEDDRDGRINOLPEREAOIRVYLEKMR BEGYTDETLIASQIIAFCGMLSRVRSSEFKYRPDDFDARPLTAOLQ"
 4469..4471
 /gene="ttk"
 /note="CC in X01714 and V01578; CAC here"
 misc_difference
 /note="CC in X01714 and V01578; CAC here"

```

4489. 4514
/note="REP (repetitive extragenic palindromic) element;
contains 1 REP sequence"
/complete(4519..5160)
/gene="pyre (CG Site No. 327)"
/complement(4519..5160)
/gene="pyre (CG Site No. 327)"
/codon_start=1
/transl_table=11
/product="orateate phosphoribosyltransferase"
/protein_id="AA61995.1"
/db_xref="GI:643610"
/translation="MKPYQRFLEALSKOVLEKFEFTLKSGKSPFPYFNGLTNGR
DLALLGGEALVADIVSEIEFDLFGAYGIPITATTLVALEHDDLDIPCFNRKKA
RDHAGEGNLVGSLACRWMLVDVITAGATATSEMTTIVQANGATLTIAGLVLSDQERG
RGESIAIOEVERDYCNCKVSIITLADLIATLEEKPEMEHLLAAVKAVREEGV"

```

Iteration	misses	hits	Indels	Gaps
1	1	0	0	0
2	1	0	0	0
3	1	0	0	0
4	1	0	0	0
5	1	0	0	0
6	1	0	0	0
7	1	0	0	0
8	1	0	0	0
9	1	0	0	0
10	1	0	0	0
11	1	0	0	0
12	1	0	0	0
13	1	0	0	0
14	1	0	0	0
15	1	0	0	0
16	1	0	0	0
17	1	0	0	0
18	1	0	0	0
19	1	0	0	0
20	1	0	0	0
21	1	0	0	0
22	1	0	0	0
23	1	0	0	0
24	1	0	0	0
25	1	0	0	0
26	1	0	0	0
27	1	0	0	0
28	1	0	0	0
29	1	0	0	0
30	1	0	0	0
31	1	0	0	0
32	1	0	0	0
33	1	0	0	0
34	1	0	0	0
35	1	0	0	0
36	1	0	0	0
37	1	0	0	0
38	1	0	0	0
39	1	0	0	0
40	1	0	0	0
41	1	0	0	0
42	1	0	0	0
43	1	0	0	0
44	1	0	0	0
45	1	0	0	0
46	1	0	0	0
47	1	0	0	0
48	1	0	0	0
49	1	0	0	0
50	1	0	0	0
51	1	0	0	0
52	1	0	0	0
53	1	0	0	0
54	1	0	0	0
55	1	0	0	0
56	1	0	0	0
57	1	0	0	0
58	1	0	0	0
59	1	0	0	0
60	1	0	0	0
61	1	0	0	0
62	1	0	0	0
63	1	0	0	0
64	1	0	0	0
65	1	0	0	0
66	1	0	0	0
67	1	0	0	0
68	1	0	0	0
69	1	0	0	0
70	1	0	0	0
71	1	0	0	0
72	1	0	0	0
73	1	0	0	0
74	1	0	0	0
75	1	0	0	0
76	1	0	0	0
77	1	0	0	0
78	1	0	0	0
79	1	0	0	0
80	1	0	0	0
81	1	0	0	0
82	1	0	0	0
83	1	0	0	0
84	1	0	0	0
85	1	0	0	0
86	1	0	0	0
87	1	0	0	0
88	1	0	0	0
89	1	0	0	0
90	1	0	0	0
91	1	0	0	0
92	1	0	0	0
93	1	0	0	0
94	1	0	0	0
95	1	0	0	0
96	1	0	0	0
97	1	0	0	0
98	1	0	0	0
99	1	0	0	0
100	1	0	0	0

57057 TATAGCGATCCCTTGCCTGAATA

57056 ACATCATCATTACGTGCGCACTGTGGCGCTATCGCACTTAAACGTTTCGTGCTGCCCT 56997

56996 CAGTCTATGCAATAGACCATAAAGTCCGCTGATAAGGCTTGAACCTT 56937

[illegible]

|||||CTGGGCGTTCAGGACCATTTGATT 300
56876 TATCCGTACCGCCTTCTTCTTCTTCTTCTT

SULT 5

DEFINITION Mus musculus clone CT7-368A6, WORKING DRAFT SEQUENCE, 85 unordered pieces

KEYWORDS: HTG; HTGS_PHASE1; HTGS_DRAFT.

Eumetazoa; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Scurionathia; Muridae

Sequencing of Mouse

TITLE
Direct Submission

-----Genome Center
Center, Joint Center for Genomic Sciences, University of California, San Diego, La Jolla, CA 92037, USA

```

-----
web site: http://www.jgl.doe.gov
-----

```

Center clone name: RG-MBAC_368A6

*	1110: contig of 1110 bp in length
*	1111: gap of unknown length
*	1211: contig of 1247 bp in length
*	2457: gap of unknown length
*	3663: contig of 1106 bp in length
2558	

	contig of	gap of	gap of
6153	1156 bp in length		
6252	unknown length		
6253	contig of 1036 bp in length		
7288	gap of unknown length		
7388	gap of unknown length		
7500			

	contig of 1312 bp in length
10275	11586: contig of 1312 bp in length
11587	11686: gap of unknown length
11687	12873: contig of 1187 bp in length
12874	13073: contig of 1187 bp in length

15842	15941: gap of unknown length
15942	17159: contig of 1218 bp in length
17160	17259: gap of unknown length
17260	18555: contig of 1205 bp in length

19843	21273: contig of 1431 bp in length
21274	21373: gap of unknown length
21374	22750: contig of 1377 bp in length
22751	22850: gap of unknown length

25639	25639: gap of unknown length
26727	26726: contig of 1088 bp in length
26827	26826: gap of unknown length
28011	28011: contig of 1185 bp in length

31153	contig of 1221 bp in length
31252	gap of unknown length
31253	contig of 1200 bp in length
32453	gap of unknown length
32552	gap of unknown length

34875	1: gap of unknown length
36052	36051: contig of 1177 bp in length
36152	36151: gap of unknown length
37590	37590: contig of 1439 bp in length

38909 39008: gap of unknown length

*	39009	40601:	gap of 1593	bp in length
*	40602	40701:	gap of unknown	length
*	40702	41926:	contig of 1225	bp in length
*	41927	42026:	gap of unknown	length
*	42027	43993:	contig of 1957	bp in length
*	43984	44083:	gap of unknown	length
*	44084	45206:	contig of 1123	bp in length
*	45207	45306:	gap of unknown	length
*	45307	46881:	contig of 1575	bp in length
*	46882	46981:	gap of unknown	length
*	46982	48056:	contig of 1075	bp in length
*	48057	48156:	gap of unknown	length
*	48157	49668:	contig of 1512	bp in length
*	49669	49768:	gap of unknown	length
*	49769	51187:	contig of 1419	bp in length
*	51188	51287:	gap of unknown	length
*	51288	52332:	contig of 1045	bp in length
*	52333	52432:	gap of unknown	length
*	52433	53755:	contig of 1322	bp in length
*	53755	53854:	gap of unknown	length
*	53855	55581:	contig of 1127	bp in length
*	55582	55681:	gap of unknown	length
*	55682	57262:	contig of 1581	bp in length
*	57263	57362:	gap of unknown	length
*	57363	59152:	contig of 1790	bp in length
*	59153	59252:	gap of unknown	length
*	59253	60732:	contig of 1480	bp in length
*	60733	60832:	gap of unknown	length
*	60833	63030:	contig of 2198	bp in length
*	63031	63130:	gap of unknown	length
*	63131	65366:	contig of 2226	bp in length
*	65367	65467:	gap of unknown	length
*	65467	67227:	contig of 1761	bp in length
*	67228	67327:	gap of unknown	length
*	67328	68506:	contig of 1119	bp in length
*	68507	68606:	gap of unknown	length
*	68607	71259:	contig of 2653	bp in length
*	71260	73135:	gap of unknown	length
*	73136	73965:	contig of 2608	bp in length
*	73966	74065:	gap of unknown	length
*	74066	76125:	contig of 2060	bp in length
*	76126	76225:	gap of unknown	length
*	76226	77297:	contig of 1072	bp in length
*	77298	77397:	gap of unknown	length
*	77398	79646:	contig of 2249	bp in length
*	79647	79746:	gap of unknown	length
*	79747	82449:	contig of 2703	bp in length
*	82450	82549:	gap of unknown	length
*	82550	85345:	contig of 2796	bp in length
*	85346	85445:	gap of unknown	length
*	85446	86817:	contig of 1372	bp in length
*	86818	86917:	gap of unknown	length
*	86919	90145:	contig of 3228	bp in length
*	90146	90245:	gap of unknown	length
*	90246	91490:	contig of 1245	bp in length
*	91491	91590:	gap of unknown	length
*	91591	94935:	contig of 3345	bp in length
*	94936	95035:	gap of unknown	length
*	95036	97845:	contig of 2810	bp in length
*	97846	97945:	gap of unknown	length
*	97946	100449:	contig of 2504	bp in length
*	100450	100549:	gap of unknown	length
*	100550	103583:	contig of 3034	bp in length
*	103584	103683:	gap of unknown	length
*	103684	107676:	contig of 3995	bp in length
*	107677	107778:	gap of unknown	length
*	107779	110534:	contig of 2756	bp in length
*	110535	110634:	gap of unknown	length
*	110635	113491:	contig of 2857	bp in length
*	113492	113591:	gap of unknown	length
*	113592	115603:	contig of 2012	bp in length
*	115604	118703:	gap of unknown	length
*	118704	118957:	contig of 3254	bp in length

Query Match	Best Local Similarity	100.0%;	Score 300;	DB 2;	Length 252977;
Matches 300;	Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;
QY 1	ATTCATCGTGTGATCGGGGCGGTGGCGCTGATGATATAGCATCCCTTCTGTAATAAATA	60			
Db 11269	ATTATCTGTTGATCGTGGGCTGTGGCTGATGATATAGCATCCCTTCTGTAATAAATA	11210			
QY 61	ACATCATCATTCATCGTCGCACATGTGGCGGCTATGCGACATTTAACTTTGCTGTCGCCCT	120			
Db 11209	ACATCATCATTCATCGTCGCACATGTGGCGGCTATGCGACATTTAACTTTGCTGTCGCCCT	11150			
QY 121	CAGCTATGCAATGAGCATTAACATGCAAAAAAAGTCCGCTGATAGGCTTGAAGAAAGTT	180			
Db 11149	CAGCTATGCAATGAGCATTAACATGCAAAAAAAGTCCGCTGATAGGCTTGAAGAAAGTT	11090			
QY 181	CATTTCCAGACCCATTTTACATCGCTAGCCGATGAGACGCGCCTGATGCTGCTGCTG	240			
Db 11089	CATTTCCAGACCCATTTTACATCGCTAGCCGATGAGACGCGCCTGATGCTGCTGCTGCTG	11030			
QY 241	TACCTGACCTGTCCTCATTTGTGGAAGGCTTACCTTCGCTGATTTGAGAGCTATTGATT	300			
Db 11029	TACCTGACCTGTCCTCATTTGTGGAAGGCTTACCTTCGCTGATTTGAGAGCTATTGATT	10970			
RESULT 6					
AC020870/c	AC020870	256373 bp	DNA	linear	HTG 17-FEB-2000
LOCUS	AC020870				
DEFINITION	Mus musculus clone RP23-302J15, LOW-PASS SEQUENCE SAMPLING.				
ACCESSION	AC020870.2	GI:6984372			
VERSION	HTG: HTGS_PHASE0.				
KEYWORDS	Mus musculus.				
SOURCE	Mus musculus.				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
REFERENCE	1 (bases 1 to 256373)				
AUTHORS	DOE Joint Genome Institute.				
TITLE	Sequencing of Mouse				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 256373)				
AUTHORS	DOE Joint Genome Institute.				
TITLE	Direct Submission				
JOURNAL	Submitted (10-JAN-2000) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA				
COMMENT	On Feb 17, 2000 this sequence version replaced gi:6686438.				
	* NOTE: This record contains 198 individual				

* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 752: contig of 752 bp in length
* gap of unknown length
* 753 1000: contig of 248 bp in length
* gap of unknown length
* 1001 1604: contig of 604 bp in length
* gap of unknown length
* 1605 2302: contig of 698 bp in length
* gap of unknown length
* 2303 2699: contig of 397 bp in length
* gap of unknown length
* 2700 3686: contig of 987 bp in length
* gap of unknown length
* 3687 4099: contig of 413 bp in length
* gap of unknown length
* 4100 4778: contig of 679 bp in length
* gap of unknown length
* 4779 5571: contig of 793 bp in length
* gap of unknown length
* 5572 5967: contig of 396 bp in length
* gap of unknown length
* 5968 6763: contig of 796 bp in length
* gap of unknown length
* 6764 7639: contig of 876 bp in length
* gap of unknown length
* 7640 8442: contig of 803 bp in length
* gap of unknown length
* 8443 9414: contig of 972 bp in length
* gap of unknown length
* 9415 10236: contig of 822 bp in length
* gap of unknown length
* 10237 10844: contig of 608 bp in length
* gap of unknown length
* 10845 11504: contig of 660 bp in length
* gap of unknown length
* 11505 12321: contig of 817 bp in length
* gap of unknown length
* 12322 12652: contig of 331 bp in length
* gap of unknown length
* 12653 14116: contig of 1464 bp in length
* gap of unknown length
* 14117 15298: contig of 1182 bp in length
* gap of unknown length
* 15299 16234: contig of 936 bp in length
* gap of unknown length
* 16235 16376: contig of 142 bp in length
* gap of unknown length
* 16377 16664: contig of 288 bp in length
* gap of unknown length
* 16665 17631: contig of 967 bp in length
* gap of unknown length
* 17632 19241: contig of 1610 bp in length
* gap of unknown length
* 19242 20453: contig of 1212 bp in length
* gap of unknown length
* 20454 20559: contig of 106 bp in length
* gap of unknown length
* 20560 21117: contig of 558 bp in length
* gap of unknown length
* 21118 21655: contig of 538 bp in length
* gap of unknown length
* 21656 21818: contig of 163 bp in length
* gap of unknown length
* 21819 22679: contig of 861 bp in length
* gap of unknown length

* 22680 23579: gap of unknown length
* 23580 24383: gap of unknown length
* 24384 25612: gap of unknown length
* 25613 25957: gap of unknown length
* 25958 27260: gap of unknown length
* 27261 28596: gap of unknown length
* 28597 29473: gap of unknown length
* 29474 30372: gap of unknown length
* 30373 31202: gap of unknown length
* 31203 32688: gap of unknown length
* 32689 33431: gap of unknown length
* 33432 33693: gap of unknown length
* 33694 34616: gap of unknown length
* 34617 35884: gap of unknown length
* 35885 36459: gap of unknown length
* 36460 36996: gap of unknown length
* 36997 37941: gap of unknown length
* 37942 38682: gap of unknown length
* 38683 39549: gap of unknown length
* 39550 40792: gap of unknown length
* 40793 42585: gap of unknown length
* 42586 44115: gap of unknown length
* 44116 45300: gap of unknown length
* 45301 46417: gap of unknown length
* 46418 47815: gap of unknown length
* 47816 49310: gap of unknown length
* 49311 50304: gap of unknown length
* 50305 50848: gap of unknown length
* 50849 51828: gap of unknown length
* 51829 52699: gap of unknown length
* 52700 53365: gap of unknown length
* 53366 54633: gap of unknown length
* 54634 55470: gap of unknown length
* 55471 56358: gap of unknown length
* 56359 56934: gap of unknown length
* 56935 57672: gap of unknown length

FEATURES	SOURCE
*	230383
*	236251: contig of 5869 bp in length
*	236252
*	236271: gap of unknown length
*	240572: contig of 4301 bp in length
*	240573
*	240592: gap of unknown length
*	242790: contig of 2197 bp in length
*	242810
*	244695: contig of unknown length
*	244696
*	244715: gap of unknown length
*	244970: contig of 2234 bp in length
*	246990
*	247870: contig of 881 bp in length
*	247871
*	247890: gap of unknown length
*	248193: contig of 303 bp in length
*	248194
*	248214: gap of unknown length
*	250122: contig of 1909 bp in length
*	250123
*	250143
*	251367: gap of unknown length
*	251368
*	251387: gap of 1225 bp in length
*	252149: gap of unknown length
*	252750: contig of 1362 bp in length
*	252770
*	253640: gap of unknown length
*	253641
*	253660: contig of 871 bp in length
*	253661
*	255032: gap of unknown length
*	255033
*	255053
*	255054
*	255055
*	257334
*	257335: gap of 2281 bp in length
*	257336
*	258409: gap of unknown length
*	259410
*	259429: contig of 2056 bp in length
*	259430
*	260459: gap of unknown length
*	260460
*	260479: contig of 1030 bp in length
*	260480
*	260829: gap of unknown length
*	260830
*	260880
*	261662
*	261681: contig of 812 bp in length
*	261682
*	262532: gap of unknown length
*	262533
*	262552: contig of 851 bp in length
*	263351: gap of unknown length
*	263352
*	263371: contig of 789 bp in length
*	263417
*	264218
*	264337: gap of unknown length
*	264338
*	265205: gap of unknown length
*	265225: contig of 968 bp in length
*	265226
*	265236
*	266270: gap of unknown length
*	266271
*	266290: contig of 1045 bp in length
*	266291
*	266749: gap of unknown length
*	266750
*	266759: gap of 459 bp in length
*	266760
*	266770
*	267718: gap of unknown length
*	267719
*	267738: gap of 949 bp in length
*	268577
*	268578
*	268596: gap of unknown length
*	270011
*	270019: contig of 838 bp in length
*	270030: gap of unknown length
*	270031
*	271654: gap of 1414 bp in length
*	271655
*	271675
*	271676
*	272545: gap of unknown length
*	272547: contig of 871 bp in length.
Location/Qualifiers	
1. 372545	

```
misc_feature      1.    30267  
                  /note="assembly_name:Config108  
misc_feature      30288    "584100  
                  /note="assembly_name:Config107  
                  clone_end:Sp6  
misc_feature      vector_side:left"  
                  58421..88864  
misc_feature      /note="assembly_name:Config106  
misc_feature      88885..109889  
misc_feature      /note="assembly_name:Config105  
                  109910..125133
```

[illegible]

Query Match	100.0%	Score 300;	DB 2:	Length 272545;
Best Local Similarity	100.0%	Pred. No. 7.3e-87;		
Matches 300;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy 1	ATTTCATCTGTTGATCGTGGTGTTGGCCGTGATGATTTTATACGATCCCTTGCGAATAA	60		
Db 257005	ATTTCATCTGTTGATCGTGGTGTTGGCCGTGATGATTTTATACGATCCCTTGCGAATAA	60		
Qy 61	ACATCATCTTACGTGCGACCTGGCGGCTATGCGACCTTAAAGCTTTCGTCGCGCCCT	120		
Db 257065	ACATCATCTTACGTGCGACCTGGCGGCTATGCGACCTTAAAGCTTTCGTCGCGCCCT	120		
Qy 121	CAGCTATGCAATAGACCATTAACATGCGAATAAAGTCCGCTATTAAGGCTTGAAAGTT	180		
Db 257125	CAGCTATGCAATAGACCATTAACATGCGAATAAAGTCCGCTATTAAGGCTTGAAAGTT	180		
Qy 181	CATTTCGAGACCCATTTTTCATGCTGTAAGCCGATGAGAGACGCGGCTATAGGGTCTCGGC	240		
Db 257185	CATTTCGAGACCCATTTTTCATGCTGTAAGCCGATGAGAGACGCGGCTATAGGGTCTCGGC	240		
Qy 241	TACCTGACCTGTCCATGTTGGAAGGCTTACATTCGCTGATTCCTGATTCACGAGACTATTGATT	300		
Db 257245	TACCTGACCTGTCCATGTTGGAAGGCTTACATTCGCTGATTCCTGATTCACGAGACTATTGATT	300		
RESULT 8	AE005600/c	12642 bp	DNA	linear
LOCUS	Escherichia coli O157:H7 EDL33	genome, contig 3 of 3,	section 219	

GSIPVFPGEPAKIKLGLAGPLIMALLIGRIGSTGLKWNPEPSANIALREGLIVE
LSVVGKSGDENVTLNGEXLSWIGALITVAVPLTIGILAMKMNITMCGHL
ASGMDPPALANLHPISGAALSYATVPLWFLIILIPOLLAVLEMSIG
/gene="lbp3"
/note="25182"
complement(6823..7257)

CDS
/gene="lbp3"
complement(6823..7257)

/function="factor: Adaptations, atypical conditions"
/note="Residues 1 to 144 of 144 are 100.00 pct identical
to residues 1 to 144 of 144 from Escherichia coli K-12
strain MG1655. B3686"
/codon_start=1
/transl_table=11
/product="heat shock protein"
/protein_id="AAG58888.1"
/db_xref="GI:12518527"

/translation="MTMRNDLSPLMRONIGFDKLANALONAGESOSFPYNIKESD
NHYRITLALGFROEDLEIQLQESTRLSVKGTPEQEEKKMLHOGMLNPPESLSTTLA
ENMEVGATFVNGLITLIRNEPIAORIAISERPALNS"
complement(7363..7776)
/gene="lbp3"
/note="25183"
complement(7363..7776)

CDS
/gene="lbp3"
complement(7363..7776)

/function="factor: Adaptations, atypical conditions"
/note="Residues 1 to 137 of 137 are 100.00 pct identical
to residues 1 to 137 of 137 from Escherichia coli K-12
Strain MG1655. B3687"
/codon_start=1
/transl_table=11
/product="heat shock protein"
/protein_id="AAG58889.1"
/db_xref="GI:12518528"

/translation="MRNFDLSPLMRONIGFDKLANALONAGESOSFPYNIKESD
ENHYRITLALGFROEDLEIQLQESTRLSVKGTPEQEEKKMLHOGMLNPPESLSTTLA
ENHYRITLALGFROEDLEIQLQESTRLSVKGTPEQEEKKMLHOGMLNPPESLSTTLA
8007..8414
/gene="yidQ"
/note="25184"
8007..8414
/gene="yidQ"
/function="orf: Unknown function"
/note="Residues 1 to 135 of 135 are 99.25 pct identical to
residues 1 to 135 of 135 from Escherichia coli K-12 strain
MG1655. B3688"
/codon_start=1
/transl_table=11
/protein_id="AAG58890.1"
/db_xref="GI:12518529"

gene

CDS

8007..8414
/gene="yidQ"
/function="orf: Unknown function"
/note="Residues 1 to 135 of 135 are 99.25 pct identical to
residues 1 to 135 of 135 from Escherichia coli K-12 strain
MG1655. B3688"
/codon_start=1
/transl_table=11
/protein_id="AAG58890.1"
/db_xref="GI:12518529"

Query Match
Best Local Similarity 99.7% Score 298.4; DB 1: Length 12642;
Matches 299; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 ATTCATCTGTTGATGCTGGTGGCTGATGAGTTATAGCATCCCTGCTGAATA 60
8076 ATTCATCTGTTGATGCTGGTGGCTGATGAGTTATAGCATCCCTGCTGAATA 60
61 ACATCATCTGTTGATGCTGGTGGCTGATGAGTTATAGCATCCCTGCTGAATA 8017
8016 ACATCATCTGTTGATGCTGGTGGCTGATGAGTTATAGCATCCCTGCTGAATA 8017
121 CAGTCATGCAATGACCAATAACTGAAAAAAGTCCGCTGATAAAGCTTGAAGAATT 180
7956 CAGTCATGCAATGACCAATAACTGAAAAAAGTCCGCTGATAAAGCTTGAAGAATT 180
181 CATTTCAGAGCCATTTTTCATGCTAGACCGATGAGAGCGCGCTGATGCTGCTGAC 240
7996 CATTTCAGAGCCATTTTTCATGCTAGACCGATGAGAGCGCGCTGATGCTGCTGAC 240

QY 241 TACCTGACCCGTCGATTTGGAAGCTTACATTCCTGCTGATTTTCAGAGCTATGATT 300
Db 7836 TACCTGACCCGTCGATTTGGAAGCTTACATTCCTGCTGATTTTCAGAGCTATGATT 7777

RESULT 9
AP002566/c 280900 bp DNA linear BCG 07-MAR-2001
LOCUS Escherichia coli O157:H7 DNA, complete genome, section 17/20.
DEFINITION AP002566 BA000007
ACCESSION AP002566.1 GI:13363930
VERSION
KEYWORDS
SOURCE
Escherichia coli O157:H7 (strain:O157:H7, sub_strain:RIMD 0509952)
DNA.
Escherichia coli O157:H7
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.

ORGANISM
REFERENCE
AUTHORS
Makino, K., Yokoyama, K., Kubota, Y., Yutsudo, C.H., Kimura, S.,
Yamamoto, K., Ishii, K., Hattori, M., Tatsuno, I., Abe, H., Iida, T.,
Sasakawa, C., and Shinagawa, H.
Complete nucleotide sequence of the prophage VT2-Sakai carrying the
verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
derived from the Sakai outbreak
Genes Genet. Syst. 74 (5), 227-239 (1999)
20198780
2 (sites)

TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS
Ohnishi, M., Murata, T., Nakayama, K., Kuhara, S., Hattori, M.,
Kurokawa, K., Yasunaga, T., Yokoyama, K., Makino, K., Shinagawa, H. and
Hayashi, T.
Comparative analysis of the whole set of RNA operons between an
enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
Escherichia coli K-12 strain MG1655
Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
20557356
3 (sites)

TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS
Yutsudo, C.H., Kurokawa, K., Ishii, K., Hattori, M., Abe, H., Iida, T.,
Yamamoto, K., Hayashi, T., Yasunaga, T., Honda, T., Sasakawa, C. and
Shinagawa, H.
Complete nucleotide sequence of the prophage VT1-Sakai carrying the
O157:H7 strain derived from the Sakai outbreak
Gene 258 (1-2), 127-139 (2000)
20564182
4 (sites)

TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS
Hayashi, T., Makino, K., Ohnishi, M., Kurokawa, K., Ishii, K.,
Yokoyama, K., Han, C.-G., Ohtsubo, E., Nakayama, K., Murata, T.,
Ogasawara, N., Yasunaga, T., Kuhara, S., Shiba, T., Hattori, M. and
Shinagawa, H.
Complete genome sequence of enterohemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12
DNA Res. 8 (1), 11-22 (2001)
21156231
5 (bases 1 to 280900)

TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS
Ohnishi, M., Kurokawa, K., Makino, K., Yasunaga, T., Shinagawa, H. and
Hayashi, T.
Direct Submission
Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
Information Research Center: 3-1, Yamadaoka, Suita, Osaka 565-0871,
Japan (E-mail: kengen-info.osaka-u.ac.jp,
URL: http://www.gen-info.osaka-u.ac.jp/
Fax: 81-6-6879-2047)
genome project.

COMMENT
FEATURES
SOURCE
Location/Qualifiers
1..280900
/organism="Escherichia coli O157:H7"
/strain="O157:H7"
/sub_strain="RIMD 0509952"
/db_xref="taxon:83334"
complement(143..616)

gene

gene
/gene="ECS4456"
/complement(143. .616)
/note="Similar to YJAI_ECOLI YJAI_ECOLI g112367245 percent
identity 98 in 157 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence-not_experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB37881.1"
/db_xref="gi:13363931"
/translation="MNRFILADATKTCIGCTCEVACVASHQENDDAALSDFEISRI
RVTKSHSTAVACHOCEDAPCANCPDASIRSHGHIFVBOPTGICGKSCMLACPG
MDEVSRKKARAIKCDLCMHRETGPACVEACPKKALQCMDEVKQVNRRLROQPV"
/complement(860. .1675)
/gene="ECS4457"
/complement(860. .1675)
/gene="ECS4457"
/note="probable transcriptional regulator, weakly similar
to hypothetical transcriptional regulator y18r (Bacillus
subtilis) g1131233061sp|P40331 percent identity 24 in 276
aa"
/codon_start=1
/evidence-not_experimental
/transl_table=11
/product="putative transcriptional regulator"
/protein_id="BAB37880.1"
/db_xref="gi:13363932"
/translation="MLELSMTPIKIKYONGGLFISRGVGRSHARRLSEWEIFEVRGL
TIOEERFLVADGESLSLAMPNROHGVEDPDLKFWLHFEKAPDDPMLHLS
VPOQKRVVDPAVALAFROPLSEQEKORHPALEFVLLLOOLFTDARIDEQADG
TLAKKAKRLIRTOYHRPLSTALARELKCXADLDGVGRVRFHLTETEAIRHQRVRE
AKLLISARSRLTEVAALCGFNDVGRFQIRKRTGLTPAMKRRYSKEHINS"
/complement(301. .3301)
/gene="ECS4458"
/complement(301. .3301)
/gene="ECS4458"
/note="probable permease, similar to hypothetical proteins
[Salmonella typhimurium] g117442781|pir1|C65167 percent
identity 37 in 444 aa, also similar to transport proteins
[putative symporters] e.g. y1cd [Escherichia coli (K-12)]
g112851421|sp|P31435|Y1CD_EC"
/codon_start=1
/evidence-not_experimental
/transl_table=11
/product="putative permease"
/protein_id="BAB37881.1"
/db_xref="gi:13363933"
/translation="MTSPITTTDIAKRTIDALSYREKIGYGLDAGCTVYTCILNN
FLTEPVDYDTPALTAICGLFIALRVDAISDPMKVADIRQSRGRPRPQWLAI
PIGIGLITFTVPDPSKGVKILANAFGYLLISGYTAIVNFCARINTMTNRHNEV
COSMFEYLCVAGFLVSGVGLPFWALPGOGNARGVGLGVGLCAIAVMPACCFW
RERVSLTSMGKFLREHLAIGLRNDOLLINLWSPLLINVENTRGGGVYFTYVLQ
SGYSLVFTMGVTFASIIIGSVIYSLRRDPYKITYYNTLLLAALAVLMLPLSPGA
VOTLMLAVILNGVILGFTPLPLSLWAFADVDGEMTKTHRRSGNMNPNLEFIKIAM
MSASGIIISLFLFVAYQGVENOTASSLGSIAAMETLPLPLHLLAGALIRCEKLNPP
MNSRTATDLRHOVP"
/complement(5282. .3312. .5282)
/gene="ECS4459"
/complement(5282. .3312. .5282)
/gene="ECS4459"
/note="unknown, similar to hypothetical protein [Thermotoga
maritima] g117452109|pir1|F72395 percent identity 37 in
635 aa"
/codon_start=1
/evidence-not_experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB37882.1"
/db_xref="gi:13363934"
/translation="MNTISEGFVDFVSPFLGQYQOVLVDVYVLYPYONDALNDRIPE
EPSALTEFRNAGIOGCEFYGVFQDSDVAKLEAVNWSLCQKDPDALEKTADEVYIT
LIAAQCEDGLNLTFTYVKAPEERMNSLAECHETLYCAGHLIEAGVAFQATGKRRLD

VYCRADLHIDSVFGEDESKLHGYPQHPELIELALMLRYEVEEPRYLALNTNYEEDRGAA
 OPHYYDOEERKGTSHMHTYGPAMPAWYKTSOAHPLAOCOTATGHALRVYVLTG
 VAILATLHSDSKRODCRLTNMNMAROLYDTTGISGSSGEAFSSPYDLPNTYAE
 SCASLTCLMEFARMLEMEGDSQYADVMERALYNTNVLGMAIDGKHPYVYDLPKRS
 LKRNLTJHYXPKIROMFEGCACCPENIRARLYTNGHLYALPREDALTYNTAGNSWED
 PVNGHGLRLRVSGNYPMOBOYV IAVESQPVARHTLALRLPWCYPOIITLNGEYVD
 IREGYLHITLREMOBEGTDLNLTLPMPVRVYNGNPLVRHVAGVVAI0RQPLVCEKDN
 GSGVYHMLTLDLAPFTTTPFGKGLFSLKTLIAPQRYVEOSNPEODPLMHYDSAPAKQ
 TOTLTFIPMSVANSRGEGRIRYNLEKHCHP"
 complement (5412..6152)
 /gene="ECS4460"
 /complement (5412..6152)
 /note="similar to YIAT_ECOLI g111790010 percent identity
 99 in 246 aa (Conserved in E.coli K-12)"
 99 in 246 aa (Conserved in E.coli K-12)"
 /codon_start=1
 /evidence=not_experimental
 /transl_table=1
 /product="putative outer membrane protein"
 /protein_id="BAB37883.1"
 /db_xref="GI:13363933"
 /translation="MLIRNTNVALFALPEMASATSELSTGCAAYANESPYGEYNT
 KAIPLISVEGDSFYRQTITGLFLLISOSEKELSLTSMWPLEFPKNDYDMOOLDK
 ROSTANAGSYAHHEHRCMTAKASAADVLNDSNGWGLSVFIRKQITGLTSLPLAG
 LYIDENFSYXXYGISESSRSGLASYSAQAAWVPVSLAKPIIGEHVYLAASGYS
 ELPEETIDSE0
 6276..7250
 /gene="ECS4461"
 6276..7250
 /note="ECS4461"
 /note="similar to YIAD_ECOLI g111790011 percent identity
 99 in 324 aa (Conserved in E.coli K-12)"
 99 in 324 aa (Conserved in E.coli K-12)"
 /codon_start=1
 /evidence=not_experimental
 /transl_table=1
 /product="putative transcriptional regulator LysR-type"
 /protein_id="BAB37884.1"
 /db_xref="GI:13363936"
 /translation="MRTLOKRYREIKTIIVIASENISHATVILGIAQAYSKILADP
 ESKGLVAYEDFTTRQMLDLEPTALPIYINDMLDRNEQLNFIADYKHEKRGATLTA
 PTGLITLVSKHIDIKIDIGDITLSIKTCLLENARYEGVEFPDDCDVLI1SAPKDS
 LSVASFITQYAVVASQRYDEKHPISRPDELHSCILIDSMILIDNAINMFVAGS
 KRVADYRKVNGVCDNQTOSALEKHNHLGITVPRDKSVOSDLODGLVPCFQHYEMW
 LDIYLFRRKEVOPRPVQYVLDEMLREIRKQILOAQSOQLRPDAQAESD"
 complement (7247..8383)
 /gene="ECS4462"
 complement (7247..8383)
 /note="ECS4462"
 /note="similar to YIAD_ECOLI g111790012 percent identity
 99 in 378 aa, similar to YIAD_ECOLI g111790024 percent
 identity 76 in 378 aa (Conserved in E.coli K-12)"
 99 in 378 aa (Conserved in E.coli K-12)"
 /codon_start=1
 /evidence=not_experimental
 /transl_table=1
 /product="putative membrane protein"
 /protein_id="BAB37885.1"
 /db_xref="GI:13363937"
 /translation="MDLLITIRYVAFAMAMEKIFKIPVKNKTIPTAALGCIYVSGLI
 LMANVNPYFPAQAKAVISIPYVPOYTGIVLEVTQKNTMLIKKGEVLEFRLDPYROR
 VDRIMADIVLEHROKRALGAEDELMANQORAKARDFEAKERYAGSQAQKYNPSS
 ERDIDVRYNVLAEKQASVSKSSAEKOKOIDSLSLVLEHSG ISLKKQILAEEVYNLE

QY 121 CAGCTATGCAATAGACCAATAGCAAAAAAGTCCGCTGATTAAGCTTGAAGGTT 180
DB 171537 CAGCTATGCAATAGACCAATAGCAAAAAAGTCCGCTGATTAAGCTTGAAGGTT 171478
QY 181 CATTTCAGACCAATTTTACATCGTACCGATGAGACGGCGCTGATGGTGTCTGGC 240
DB 171477 CATTTCAGACCAATTTTACATCGTACCGATGAGACGGCGCTGATGGTGTCTGGC 240
QY 241 TACCTGACCTGTCATTTGAGAGTCTTACATTTCTGCTGATTTAGAGATTTGATT 300
DB 171417 TACCTGACCTGTCATTTGAGAGTCTTACATTTCTGCTGATTTAGAGATTTGATT 171358
RESULT 10 AC023611 221285 bp DNA linear HTG 29-JUN-2000
LOCUS AC023611
DEFINITION Mus musculus clone CT7-9K21, WORKING DRAFT SEQUENCE, 62 unordered
pieces.
ACCESSION AC023611 GI:8810277
VERSION AC023611.2
KEYWORDS HTG, HTGS, PHASE1, HTGS, DRAFT.
SOURCE Mus musculus.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 221285)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Mouse
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 221285)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (16-FEB-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT -----Genome Center
On Jun 29, 2000 this sequence version replaced gi:6980221.
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov

Project Information
Center Project Name: 956898
Center Clone Name: RG-MBAC_9K21

Summary Statistics
Consensus quality: 172196 bases at least Q40
Consensus quality: 196011 bases at least Q30
Consensus quality: 201702 bases at least Q20
Estimated insert size: 160000.
Estimated insert size: 215185; pulse field gel estimation
Quality coverage: 6.26 in Q20 bases; pulse field gel estimation
Quality coverage: 4.65 in Q20 bases; sum-of-coverage estimation
NOTE: This is a 'working draft' sequence. It currently
* consists of 62 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
1034 1033: contig of 1033 bp in length
1134 1133: gap of unknown length
2542 2542: contig of 1409 bp in length
2643 2642: gap of unknown length
3698 3698: contig of 1056 bp in length
3799 3798: gap of unknown length
4993 4993: contig of 1195 bp in length
4994 5093: gap of unknown length
5094 6427: contig of 1334 bp in length
6428 6527: gap of unknown length
7763 7763: contig of 1236 bp in length
7764 7863: gap of unknown length

8877: contig of 1014 bp in length
8978 8978: gap of unknown length
10256 10256: contig of 1279 bp in length
10356 10356: gap of unknown length
11356 11356: contig of 1000 bp in length
11457 11457: gap of unknown length
12471 12470: contig of 1014 bp in length
12571 12571: gap of unknown length
13862 13861: contig of 1291 bp in length
13962 13961: gap of unknown length
15061 15060: contig of 1099 bp in length
15161 15160: gap of unknown length
16252 16252: contig of 1092 bp in length
16353 16352: gap of unknown length
17471 17470: contig of 1118 bp in length
17571 17570: gap of unknown length
18938 18937: contig of 1267 bp in length
20247 20246: gap of unknown length
20347 20346: contig of 1309 bp in length
21516 21515: gap of unknown length
21616 21615: contig of 1169 bp in length
22753 22752: contig of 1137 bp in length
22853 22852: gap of unknown length
24439 24438: contig of 1586 bp in length
24539 24538: gap of unknown length
25616 25615: contig of 1077 bp in length
25716 25715: gap of unknown length
26919 26918: contig of 1203 bp in length
27019 27018: gap of unknown length
28162 28161: contig of 1143 bp in length
28262 28261: gap of unknown length
29907 29906: contig of 1645 bp in length
30007 30006: gap of unknown length
31336 31335: contig of 1329 bp in length
31436 31435: gap of unknown length
32662 32661: contig of 1226 bp in length
32762 32761: gap of unknown length
34093 34093: contig of 1332 bp in length
34194 34193: gap of unknown length
35483 35482: contig of 1289 bp in length
35582 35582: gap of unknown length
37002 37001: contig of 1419 bp in length
37102 37101: gap of unknown length
38337 38337: contig of 1236 bp in length
38437 38437: gap of unknown length
38438 38438: gap of unknown length
39825 39825: contig of 1388 bp in length
39925 39925: gap of unknown length
41162 41162: gap of unknown length
41262 41262: contig of 1237 bp in length
42872 42872: contig of 1610 bp in length
42972 42972: gap of unknown length
44118 44118: contig of 1146 bp in length
44218 44218: gap of unknown length
45913 45913: contig of 1693 bp in length
46013 46013: gap of unknown length
47295 47295: contig of 1282 bp in length
47396 47396: gap of unknown length
48000 48000: contig of 1105 bp in length
48501 48501: gap of unknown length
50670 50670: gap of unknown length
50770 50770: contig of 2070 bp in length
52744 52744: gap of unknown length
52845 52844: contig of 1974 bp in length
54280 54280: gap of unknown length
54381 54380: contig of 1436 bp in length
56311 56311: gap of unknown length
56411 56411: contig of 1931 bp in length
59166 59166: contig of 2755 bp in length
59266 59266: gap of unknown length
62681 62681: contig of 3415 bp in length
62781 62781: gap of unknown length
65691 65691: contig of 2910 bp in length

[illegible]

```

FEATURES
source
    Location=1. 245050
    /organism="Salmonella enterica subsp. enterica serovar Typhi"
    /strain="C118"
    /db_xref="taxon:90370"
    319..1728
    /gene="STY3874"
    /note="g1na"
    319..1728
    /gene="STY3874"
    /note="Orthologue of E. coli g1na (GLNA_ECOLI); FastA hit to GLNA_ECOLI (468 aa). 98% identity in 468 aa overlap"

```

```

/codon_start=1
/transl_table=11
/product="glutamine synthetase"
/protein_id="CAD03093.1"

```

```

/misc_feature
/gene="STY3874"
/feature="Pfam match to entry PF00120 gln-synt, Glutamine
synthetase, score 838.00, E-value 2.4e-266"
466..522
/gene="STY3874"
/note="PS00180 Glutamine synthetase signature 1"
1093..1140
/gene="STY3874"
misc_feature

```



```

Db 110272 ACCTATTACAGACCATTTATAGACGTAGTGAAGAGGACCCGATGGGTCT 110331
OY 238 GGCTA-CCGTACCTGTCATGTCAGAGTCTTACATCTGCTATTCAGAGCTANT 296
Db 110332 GGCAATCCGCGCTTCATGATGAGAGCTGATGATCTGCTATTCAGAGTATT 110391
OY 297 GATT 300
Db 110392 GATT 110395

RESULT 12
AE008877 22492 bp DNA linear BCT 31-JUL-2002
LOCUS Salmoneella typhimurium LT2, section 181 of 220 of the complete
DEFINITION genome.
ACCESSION AE008877 AE006468
VERSION AE008877.1 GI:16422367
KEYWORDS
SOURCE
ORGANISM Salmoneella typhimurium LT2.
REFERENCE 1 (bases 1 to 22492)
AUTHORS Mclelland, M., Sanderson, K. E., Spieth, J., Clifton, S. W.,
Latreille, P., Courtney, L., Portolillo, S., All, J., Dante, M., Du, F.,
Hou, S., Layman, D., Leonard, S., Nguyen, C., Scott, K., Holmes, A.,
Grewal, N., Mulvaney, E., Ryan, E., Sun, H., Florea, L., Miller, W.,
Stonkeling, T., Nhan, M., Waterston, R. and Wilson, R. K.
Complete genome sequence of Salmoneella enterica serovar Typhimurium
LT2
JOURNAL Nature 413 (6858), 852-856 (2001)
MEDLINE 21354948
PUBMED 11677609
2 (bases 1 to 22492)
REFERENCE The Salmoneella typhimurium Genome Sequencing Project.
AUTHORS Direct Submission
JOURNAL Submitted (29-MAR-2001) Genome Sequencing Center, Department of
Genetics, Washington University School of Medicine, 4444 Forest
Park Boulevard, St. Louis, MO 63108, USA
COMMENT Supported by NIH Grant 5U 01 AI43283

COMMENT
Coding sequences below are predicted from manually evaluated
computer analysis, using similarity information and the programs;
GLIMMER, http://www.tigr.org/softlab/glimmer/glimmer.html and
GeneMark; http://opal.biology.gatech.edu/GeneMark/
EC numbers were kindly provided by Junko Yabuzaki and the Kyoto
Encyclopedia of Genes and Genomes; http://www.genome.ad.jp/kegg/,
and Pedro Romero and Peter Karp at EcoCyc;
http://ecocyc.org/ecocyc/

The analyses of ribosome binding sites and promoter binding sites
were kindly provided by Heladia Salgado, Julio Collado-Vides and
Reguondb;
http://kinich.cifn.unam.mx:8850/db/reguondb\_intro.frameset

This sequence was finished as follows unless otherwise noted: all
regions were double stranded, sequenced with an alternate
chemistries or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one m33 subclone.

FEATURES
source
1..22492
/organism="Salmoneella typhimurium LT2"
/strain="LT2; SGSC 1412; ATCC 700720"
/db_xref="ATCC:700720"
/db_xref="taxon:99287"
/note="LT2"
complement(78..278)
/gene="lvbL"

CDS
10..22492
/note="STM3797"
complement(78..176)
/gene="lvbL"
/note="similar to E. coli lvb operon leader peptide
(AAC76695.1); Blastp hit to AAC76695.1 (32 aa), 84%
identity in aa 1 - 32"
/codon_start=1
/transl_table=1
/product="lvb operon leader peptide"
/protein_id="AAL22655.1"
/db_xref="GI:16422368"
/translation="MNPSTLNTLITLTPASRAVVVVVVVGNAP"
complement(181..186)
/gene="lvbL"
/note="putative RBS for lvbL; Reguondb:STM3H003766"
complement(254..262)
/gene="lvbL"
/note="putative -10 signal for lvbL;
Reguondb:STM3H004613"
complement(264..292)
/note="putative binding site for CRP, Reguondb:
STM3H000038"
/bound_molety="CRP"
complement(270..278)
/gene="lvbL"
/note="putative -35 signal for lvbL;
Reguondb:STM3H004613"
693..813
/gene="ysdA"
/note="STM3797A"
join(699..790,789..813)
/gene="ysdA"
/note="ribosomal slippage"
/codon_start=1
/transl_table=1
/product="ysdA"
/protein_id="AAM89474.1"
/db_xref="GI:22025623"
/translation="MSVVDITLILKILVALQLDAVLYLK"
810..899
/gene="ysdB"
/note="STM3796B"
810..899
/gene="ysdB"
/codon_start=1
/transl_table=1
/product="conserved protein in the LexA regulon"
/protein_id="AAL22656.1"
/db_xref="GI:16422369"
/translation="MSVVDITLILKILVALQLDAVLYLK"
922..1852
/gene="STM3796A"
922..1852
/gene="STM3796A"
/codon_start=1
/transl_table=1
/product="integral membrane protein"
/protein_id="AAL22657.1"
/db_xref="GI:16422370"
/translation="WTPSTPDAMTISVFCILPFAALLHASMNAIVKAGNDKTLAAG
VSGSAVVALILPSPPOPAHSIDPFAASALGVYTVLVAKYVSDMSQYPLMR
GTAFLVLAIVSLVGLDLSLAWGVAVICMALLIGMAACNRASRCYVATLNACF
IAGYLVGTGVSERLAVTWSFUNGACLTWMAIARRRARSRLAQQKKGIF
AGIGTMSYGLALNMTQAPLAVVAALRETSILFGALIAWLLEKVAIGLVAAGI
ALGAILRLS"
2055..3239
/gene="emrD"
/note="STM3798"
2055..3239
/gene="emrD"
/note="similar to E. coli 2-module integral membrane pump;
multidrug resistance (AAC76696.1); Blastp hit to
AAC76696.1 (396 aa), 92% identity in aa 3 - 396"

```


STS 176049. .176250
/note="matching EMBL:T89997

RHdb:RH26726

dbSTS:STS19237

Identified using the e-PCR software (G. Schuler)"

187792. .187892

/note="matching EMBL:G20753

RHdb:RH17210

dbSTS:STS20239

Identified using the e-PCR software (G. Schuler)"

BASE COUNT 59443 a 40899 c 40806 g 53857 t 2 others

ORIGIN

Query Match

Best Local Similarity 12.1%; Score 36.2; DB 9; Length 195007;

Matches 65; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

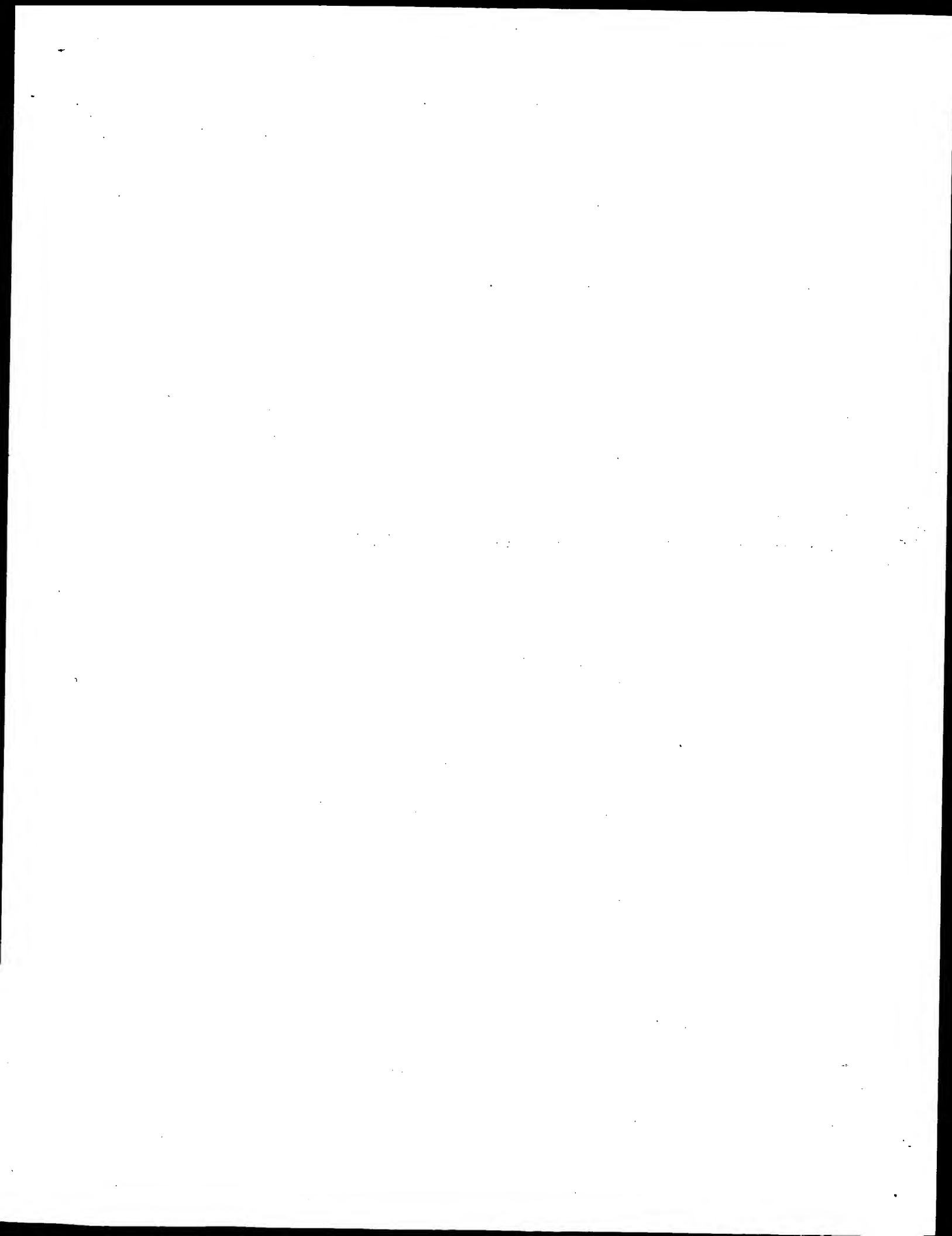
QY 106 TTGCTGCTGCCCCCTGAGTGTATGCAATGACCATTAAGTCAAAAAAGTCCGCTGAT 165

Db 138137 TTGCTGCTGAGGAGGAGATTGAGATGACATGACAAACGCAAAATTAACCCACCT 138078

QY 166 AAGGCTTGAAAGTTCATTTCCAGACCATTTTACATCGTAGCCGATGAGGA 218

Db 138077 GTTCTTGATTTTAATTTCAACAGACTTTACATTTCCACTGCGGAGGA 138025

Search completed: May 5, 2003, 19:21:18
Job time: 2053.7 secs



GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd

OM nucleic - nucleic search, using SW model

Run on: May 5, 2003, 17:50:42 ; Search time 237.363 Seconds
(without alignments)
2846.276 Million cell updates/sec

```

Title:      US-09-990-099-21
Perfect score: 300
Sequence:   1 atcatctgttgatcgtggy.....gattcagagctatgtatt 300

```

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.00

```
Searched: 2185239 seqs, 1125999159 residues
Total number of hits satisfying chosen parameters: 4370478
```

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

```

Post-processing:  Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries

```

Database :

1:	/N/genseq/9cgcdata./genseq/gensesqn_emb1/NAL1980.DAT.*
2:	/SIDS2/gcgdata./genseq/gensesqn_emb1/NAL1981.DAT.*
3:	/SIDS2/gcgdata./genseq/gensesqn_emb1/NAL1982.DAT.*
4:	/SIDS2/gcgdata./genseq/gensesqn_emb1/NAL1983.DAT.*
5:	/SIDS2/gcgdata./genseq/gensesqn_emb1/NAL1984.DAT.*
6:	/SIDS2/gcgdata./genseq/gensesqn_emb1/NAL1985.DAT.*
7:	/SIDS2/gcgdata./genseq/gensesqn_emb1/NAL1986.DAT.*
8:	/SIDS2/gcgdata./genseq/gensesqn_emb1/NAL1987.DAT.*
9:	/SIDS2/gcgdata./genseq/gensesqn_emb1/NAL1988.DAT.*
10:	/SIDS2/gcgdata./genseq/gensesqn_emb1/NAL1989.DAT.*
11:	/SIDS2/gcgdata./genseq/gensesqn_emb1/NAL1990.DAT.*
12:	/SIDS2/gcgdata./genseq/gensesqn_emb1/NAL1991.DAT.*
13:	/SIDS2/gcgdata./genseq/gensesqn_emb1/NAL1992.DAT.*
14:	/SIDS2/gcgdata./genseq/gensesqn_emb1/NAL1993.DAT.*
15:	/SIDS2/gcgdata./genseq/gensesqn_emb1/NAL1994.DAT.*
16:	/SIDS2/gcgdata./genseq/gensesqn_emb1/NAL1995.DAT.*
17:	/SIDS2/gcgdata./genseq/gensesqn_emb1/NAL1996.DAT.*
18:	/SIDS2/gcgdata./genseq/gensesqn_emb1/NAL1997.DAT.*
19:	/SIDS2/gcgdata./genseq/gensesqn_emb1/NAL1998.DAT.*
20:	/SIDS2/gcgdata./genseq/gensesqn_emb1/NAL1999.DAT.*
21:	/SIDS2/gcgdata./genseq/gensesqn_emb1/NAL2000.DAT.*
22:	/SIDS2/gcgdata./genseq/gensesqn_emb1/NAL2001A.DAT.*
23:	/SIDS2/gcgdata./genseq/gensesqn_emb1/NAL2001B.DAT.*
24:	/SIDS2/gcgdata./genseq/gensesqn_emb1/NAL2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result	Score	Query	Match	length	DB	ID	Description
No.							
1	236.4	78.8	815	23	AAS93669	DNA encoding novel	
2	34.4	11.5	866	22	AAH03353	Human cDNA clone (
3	34.4	11.5	1791	21	AAC98957	Human pancreatic c	
4	34.4	11.5	1853	22	AAH13701	Human cDNA sequen	
5	34.4	11.5	2162	21	AACT6764	Human OREF ORF2319	
6	34.4	11.5	2318	24	ABG89998	Human poly(ADP-ribo	
7	32	10.7	720	22	AAI21446	Human breast cancer	
c	30.8	10.3	765	23	ABG84698	DNA encoding novel	
8	30.6	10.2	1252	24	ABL49512	Sequence #114 used	
9							

C 10	30.6	10.2	1252	24	ABK30078	Plant dwarfing/stu
C 11	30.6	10.2	1584	21	AARD00097	Mirine zsig49 CDNA
C 12	30.4	10.1	443	21	AAC25833	Human secreted pro
C 13	30.4	10.1	568	22	AAS29929	Human lung antigen
C 14	30.4	10.1	1579	22	ABA21158	Human nervous syst
C 15	30.4	10.1	1579	22	AAS30161	Human lung antigen
C 16	30.4	10.1	4607	22	ABA21157	Human nervous syst
C 17	30.4	10.1	4607	22	AAS30158	Human lung antigen
C 18	30.4	10.1	4607	22	AAS30160	Human immune syste
C 19	30.4	10.1	9965	22	ABJ33526	Human immune syste
C 20	30.4	10.1	163319	21	AAE22306	Arabidopsis thalia
C 21	30.2	10.1	1579	22	ABA21156	Human nervous syst
C 22	30.2	10.1	1579	22	AAS30159	Human lung antigen
C 23	30	10.0	8456	22	ABL33476	Human immune syste
C 24	29.8	9.9	2437	19	AAV55887	Human lung antigen
C 25	29.8	9.9	1917	23	AAV65309	Human immune syste
C 26	29.6	9.9	1917	23	AA655868	Human lung antigen
C 27	29.6	9.9	6000	24	ABN9975	Human immune syste
C 28	29.6	9.9	10024	24	ABN95933	Receptor protein t
C 29	29.4	9.8	1701	20	AAV83626	DNA encoding novel
C 30	29.4	9.8	24025	17	AAAT17455	Novel human coding
C 31	29.4	9.8	24025	17	AAAT17515	Gene #2431 used to
C 32	29.4	9.8	24026	17	AAAT17512	Nucleic acid encod
C 33	29.4	9.8	24026	17	AAAT17513	Mutated BRCA1 geno
C 34	29.4	9.8	24026	17	AAAT17516	Mutated BRCA1 geno
C 35	29.4	9.8	24026	17	AAAT17516	Mutated BRCA1 geno
C 36	29.4	9.8	24026	17	AAAT17517	Mutated BRCA1 geno
C 37	29.4	9.8	24026	17	AAAT17518	Mutated BRCA1 geno
C 38	29.4	9.8	24026	17	AAAT17519	Mutated BRCA1 geno
C 39	29.4	9.8	24026	17	AAAT17522	Mutated BRCA1 geno
C 40	29.4	9.8	24026	17	AAAT17523	Mutated BRCA1 geno
C 41	29.4	9.8	24026	17	AAAT17524	Mutated BRCA1 geno
C 42	29.4	9.8	24026	17	AAAT17525	Mutated BRCA1 geno
C 43	29.4	9.8	24026	17	AAAT17526	Mutated BRCA1 geno
C 44	29.4	9.8	24026	17	AAAT17527	Mutated BRCA1 geno
C 45	29.4	9.8	24026	17	AAAT17528	Mutated BRCA1 geno

ALIGNMENTS

RESULT 1
AAS93669
ID AAS93669 standard; CDNA; 815 BP.

DT 13-FEB-2002 (first entry)

AA DNA encoding novel human diagnostic protein #29473.
DE

xx Human; chromosome mapping; gene mapping; gene therapy; forensic;
 kw food supplement; medical imaging; diagnostic; genetic disorder; ss

OS Homo sapiens.

PN W0200175067-A2

PD 11-OCT-2001

PF 30-MAR-2001; 2001

PR 31-MAR-2000; 2000US-0540217

XX

XX
XX
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
8

XX 2001-630363/73

DR P-PSDB; ABG29482.

PT New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity

Claim 1; SEQ ID NO 29473; 103pp; English.

The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations in
 CC responsible for genetic disorders or other traits to assess biodiversity,
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

Sequence 815 BP: 210 A; 203 C; 187 G; 215 T; 0 other;

Query Match
 Best Local Similarity 78.8%; Score 236.4; DB 23; Length 815;
 Matches 240; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

YY 55 AAAAAATCATCATCATGACGCTGCTGGCGCTATCGCAGCTTACGTTTCGCTG 114
 DB 8 AAAAAATCATCATCATGACGCTGCTGGCGCTATCGCAGCTTACGTTTCGCTG 114
 YY 115 CCCCCAGTCTATGCAATGACCAATGCAAAAAAGTCCGCTGTAAGGCTGA 174
 DB 68 CCCCCAGTCTATGCAATGACCAATGCAAAAAAGTCCGCTGTAAGGCTGA 174
 YY 175 AAGTCTATGCTGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 234
 DB 128 AAGTCTATGCTGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 187
 YY 235 TCGGCTACCTGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 294
 DB 188 TCGGCTACCTGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 247
 YY 295 TTGATT 300
 DB 248 TTGATT 253

RESULT 2
 AAH03353 standard; cDNA; 866 BP.

AAH03353;

26-JUN-2001 (first entry)

Human cDNA clone (5'-primer) SEQ ID NO:188.

Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

OS Homo sapiens.
 XX EPI074617-A2.
 XX 07-FEB-2001.
 XX
 XX
 XX

DE Human pancreatic cancer antigen nucleotide sequence SEQ ID NO:225.

PF 28-JUL-2000; 2000EP-0116126.
 XX
 PR 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 DR WPI: 2001-318749/34.

PT primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -

PS Claim 1; SEQ ID 188; 2537pp + CD ROM; English.

CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC of an oligonucleotide comprises a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human amino acid sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.

Sequence 866 BP: 219 A; 228 C; 215 G; 201 T; 3 other;

Query Match
 Best Local Similarity 11.5%; Score 34.4; DB 22; Length 866;
 Matches 71; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

YY 153 AAGTCCGCTGATTAAGGCTTGAAGTTCATTTCCAGACCCATTTTACATGCGTAGCGGA 212
 DB 353 AAGGCTGCTGTTGCTTGCATGACAGATCANTTTATTCACACATTTAAAGACATGAAGC 412
 YY 213 TGAGAGCGCGCTGATGAGGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 272
 DB 413 TGTGAGAGCGCTGATGAGGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 472
 YY 273 TTCTGCTGATT 284
 DB 473 ACCATTTCCAAAT 484

RESULT 3
 AAC8997 standard; cDNA; 1791 BP.

AAC8997;

09-MAR-2001 (first entry)

DE Human pancreatic cancer antigen nucleotide sequence SEQ ID NO:225.

XX Human; pancreas; pancreatic cancer; pancreatic cancer antigen;
 KW detection; diagnosis; identification; cytostatic; neuroprotective;
 KW neoplastic; immunomodulatory; relaxant; contraceptive; gynaecological;
 KW antiinflammatory; cardiant; gene therapy; chromosome mapping;
 KW linkage analysis; tissue identification; tissue typing; forensic;
 KW neural; immune system; muscular; reproductive; gastrointestinal;
 KW pulmonary; cardiovascular; renal; proliferative; ss.
 OS Homo sapiens.
 XX WO200055320-A1.
 XX 21-SEP-2000.
 XX 08-MAR-2000; 2000WO-US05989.
 XX 12-MAR-1999; 99US-0124270.
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX Rosen CA, Ruben SM;
 XX WPI: 2000-579444/54.
 XX P-PDB: AAB54232.
 XX New nucleic acid that is a pancreatic cancer antigen for preventing,
 PT treating, or ameliorating a medical condition, particular pancreatic
 PT cancer, or for use in assays for diagnosing a pathological condition -
 PS Claim 1; Page 668-669; 1379pp; English.
 XX AAC98773 to AAC99231 encode the human pancreatic cancer associated
 CC proteins, called pancreatic cancer antigens, given in AAB54008 to
 CC AAB54466. The human pancreatic cancer antigens have cytostatic,
 CC neuroprotective, neoplastic, immunomodulatory, relaxant, contraceptive,
 CC gynaecological, cardiant and antiinflammatory activities, and can be used
 CC in gene therapy. The polynucleotide and proteins can be used for
 CC preventing, treating, or ameliorating a medical condition or in assays
 CC for diagnosing a pathological condition or a susceptibility to one in a
 CC subject. Binding partners to the proteins and the activity of the
 CC proteins can be identified. The pancreatic cancer antigens can be used to
 CC detect, treat or prevent pancreatic disorders, especially cancer.
 CC Agonists and antagonists to the antigens can be screened for. The
 CC pancreatic cancer antigen polynucleotides can be used to design nucleic
 CC acid hybridisation probes that can be used in chromosome mapping, linkage
 CC analysis, tissue identification and/or typing and a variety of forensic
 CC and diagnostic methods. The proteins can be used to generate antibodies
 CC which are used to purify, detect and target the polypeptides, including
 CC both in vivo and in vitro diagnostic and therapeutic methods. The
 CC proteins can be used to treat or prevent neural, immune system, muscular,
 CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
 CC proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
 CC sequences used in the exemplification of the present invention.
 XX SO Sequence 1791 BP; 497 A; 422 C; 420 G; 452 T; 0 other;
 Query Match 11.5%; Score 34.4; DB 21; Length 1791;
 Best Local Similarity 53.8%; Pred. No. 0.19;
 Matches 71; Conservative 0; Mismatches 61; Indels 0; Gaps 0;
 OY 153 AAGTCCGCTGATAGCTTGAAGAGTTCATTCAGACCCATTTCATTCAGTACCGCA 212
 DB 276 AAGGCTGCTGTTGCTTGCCTAGAGACTCATTTATTTACACATTAAGACATGAAGC 335
 OY 213 TGAGAGAGCCCTGATGGGTTCTGCTACCTGACCTGCTCATTTGTGGAAGCTCTTACA 272
 DB 336 TGTGAAGACCTCTCTGATATTCCTGCAACCAACAGGTATGTGCTCTCTATCA 395
 OY 273 TTCTGCTGATT 284
 DB 396 ACCATTCATTT 407

RESULT 4
 AAH13701
 ID AAH13701 standard; cDNA; 1853 BP.
 XX AAH13701;
 AC 26-JUN-2001 (first entry)
 XX DE Human cDNA sequence SEQ ID NO:10576.
 XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
 XX Homo sapiens.
 XX EP1074617-A2.
 XX 07-FEB-2001.
 XX 28-JUL-2000; 2000EP-0116126.
 XX 29-JUL-1999; 99JP-0248036.
 XX 27-AUG-1999; 99JP-0300253.
 XX 11-JAN-2000; 2000JP-0118776.
 XX 02-MAY-2000; 2000JP-0183767.
 XX 09-JUN-2000; 2000JP-0241899.
 XX (HELI-) HELIX RES INST.
 XX Oca T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PT Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 PI WPI: 2001-318749/34.
 XX Claim 8; SEQ ID 10576; 2537pp + CD ROM; English.
 XX The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB52446 to AAH13632
 CC AAB59893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX SO Sequence 1853 BP; 503 A; 448 C; 445 G; 457 T; 0 other;
 Query Match 11.5%; Score 34.4; DB 22; Length 1853;
 Best Local Similarity 53.8%; Pred. No. 0.19;
 Matches 71; Conservative 0; Mismatches 61; Indels 0; Gaps 0;
 OY 153 AAGTCCGCTGATAGCTTGAAGAGTTCATTCAGACCCATTTCATTCAGTACCGCA 212
 DB 353 AAGGCTGCTGTTGCTTGCCTAGAGACTCATTTATTTACACATTAAGACATGAAGC 412

QY 213 TGAAGACGGCCGTGATGGTGTCTGCTACCGACCTTCATTTGCGAGGCTTACA 272
 DB 413 TGTGTAAGACCCCTCGATATTCCTCGCAACCAACAGCTGTATGTCTCTCTACCA 472
 OY 273 TTCTGCGCTATT 284
 DB 473 ACCATTCCCAATT 484

RESULT 5

AAC76764
 ID AAC76764 standard; cDNA: 2162 BP.

AC 6764;

DT 08-FEB-2001 (first entry)

DE Human ORF2319 polynucleotide sequence SEQ ID NO:4637.

XX Human; open reading frame; ORF; detection; cytosolic; hepatotropic;
 KW vulnary; antiparasitic; antiparkinsonian; neurotropic; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiac;
 KW hypotensive; dermatological; coagulant; vasotrophic; antidiabetic;
 KW antiviral; antibacterial; antifungal; antirheumatic; antihydroid;
 KW antineoplastic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypochyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KW thrombosis; contraceptive; ss.

XX Homo sapiens.

OS WO200058473-A2.

XX 05-OCT-2000.

PD 31-MAR-2000; 2000WO-US08621.

PF 31-MAR-1999; 99US-0127607.

PR 02-APR-1999; 98US-0127636.

PR 05-APR-1999; 98US-0127728.

PR 30-MAR-2000; 2000US-0340763.

XX (CUBA-) CURAGEN CORP.

PI Shimkels RA, Leach M;

XX WPI: 2000-602362/57.

DR P-PSDB; ABB42555.

XX Novel nucleic acids and peptides derived from open reading frame X,

PT useful for treating e.g. cancers, proliferative disorders,

XX neurodegenerative disorders and cardiovascular disease.

PS Claim 5; Page 3835-3836; 5507pp; English.

XX AAC74446 to AAC77606 encode the proteins given in ABB40237 to ABB43397,
 CC which represent the human ORF open reading frames 1 to 3161. The ORF
 CC sequences have activities such as: cytosolic; hepatotropic; vulnary;
 CC antiparasitic; antiparkinsonian; neurotropic; neuroprotective;
 CC osteopathic; anticonvulsant; antidiabetic; immunosuppressant;
 CC immunostimulant; cardiac; thrombolytic; coagulant; vasotrophic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antihydroid; antineoplastic; antiviral; antifungal; antirheumatic;
 CC antihydroid; antineoplastic; antiviral; antifungal; antirheumatic;
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORF-associated disorder. The
 CC nucleic acids can be used to express ORF proteins in gene therapy.

CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypochyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.

SO Sequence 2162 BP; 575 A; 511 C; 526 G; 548 T; 2 other:

Query Match 11.5%; Score 34.4; DB 21; Length 2162;

Best Local Similarity 53.8%; Pred. No. 0.21; Mismatches 61; Indels 0; Gaps 0;

Matches 71; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 153 AAGTCCGCTGATTAAGCTTGAAGAGTTCATTTCCAGACCAATTTTACATGCTAGCCGA 212
 DB 401 AAGGCTGCTGGTTTCTGCTAGAGAGCTCATTTATTCACAACTTAAGACATGAAC 460
 QY 213 TGAAGACGGCCGTGATGGTGTCTGCTACCGACCTTCATTTGCGAGGCTTACA 272
 DB 461 TGTGTAAGACCCCTCGATATTCCTCGCAACCAACAGCTGTATGTCTCTCTACCA 520
 OY 273 TTCTGCGCTATT 284
 DB 521 ACCATTCCCAATT 532

RESULT 6

ABL89998
 ID ABL89998 standard; cDNA: 2318 BP.

XX ABL89998;

DT 24-MAY-2002 (first entry)

XX Human polynucleotide seq ID NO 560.

XX Cytosolic; immunosuppressive; neurotropic; neuroprotective; antiviral;
 KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antileuc;
 KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;
 KW cardiac; gene therapy; cancer; immune disorder; cardiovascular disorder;
 KW neurological disease; infection; human; secreted protein; gene; ss.

XX Homo sapiens.

OS WO200190304-A2.

XX 29-NOV-2001.

PD 18-MAY-2001; 2001WO-US16450.

PF 19-MAY-2000; 2000US-205515P.

PR (HUMA-) HUMAN GENOME SCI INC.

PI Birse CE, Rosen CA;

XX WPI: 2002-122018/16.

DR P-PSDB; ABB89589.

XX Novel 1405 isolated polypeptides, useful for diagnosis, treatment and

PT prevention of neutral, immune system, muscular, reproductive,

XX gastrointestinal, pulmonary, cardiovascular, renal and proliferative

PS Claim 4; SEQ ID NO 560; 2081pp + Sequence Listing; English.

XX The invention relates to novel genes (ABL89449-ABL90853) and proteins
 CC (ABB89440-ABB90444), useful for preventing, treating or ameliorating
 CC medical conditions e.g. by protein or gene therapy. The genes are
 CC isolated from a range of human tissues disclosed in the specification.

CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemia and
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pat_sequences.

XX Sequence 2318 BP: 629 A; 538 C; 569 G; 582 T; 0 other;

Query Match 11.5%; Score 34.4; DB 24; Length 2318;
Best Local Similarity 53.8%; Pred. No. 0.22; Mismatches 0; Gaps 0;
Matches 71; Conservative 0; Indels 0;

OY 153 AAGTCCGCTGATAGGCTGAAAGTTCATTCCAGACCCATTTTTCATCGTAGCCGA 212
DB 436 AAGGCTGCTGTTGCTGCTAGAGAGTCATTTATTCACAAATTAAGACATGAAGC 495
OY 213 TGAGAGCGCGCTGATGGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 272
DB 496 TGTGAGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 555
OY 273 TGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 284
DB 556 ACCATTCGCAATT 567

RESULT 7

AA121446/C
ID AA121446 standard; cDNA: 720 BP.

XX AA121446;

DT 07-DEC-2001 (first entry)

DE Human breast cancer expressed polynucleotide 13903.

KW Human; breast cancer; cell marker; cytosstatic; ss.

OS Homo sapiens.

PN WO200151628-A2.

PD 19-JUL-2001.

PF 10-JAN-2001; 2001WO-US00798.

PR 14-JAN-2000; 2000US-0176077.

PR 14-MAR-2000; 2000US-0189167.

PR 24-MAR-2000; 2000US-0192099.

PR 29-MAR-2000; 2000US-0193480.

PR 15-MAY-2000; 2000US-0205230.

PR 09-JUN-2000; 2000US-0211315.

PR 25-JUL-2000; 2000US-0220534.

XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX Lillie J, Xu Y, Wang Y, Steinmann K.

XX WPI: 2001-451856/48.

XX New peptide useful as a marker for the diagnosis of breast cancer
XX Claim 1; Page 2476; 3695pp; English.
XX The invention relates to human breast cancer expressed polynucleotides

CC (AA107544-AA126789) and methods of assessing whether a patient is
CC afflicted with breast cancer by examining the correlation between the
CC expression of certain markers and the cancerous state of breast cells.
CC The polynucleotides and encoded polypeptides are potential markers for
CC detecting, diagnosing, monitoring, characterising treating and
CC potentially preventing breast cancer. The polynucleotides and encoded
CC polypeptides are also useful for isolating compounds with cytosstatic
CC activity.

XX Sequence 720 BP: 198 A; 163 C; 115 G; 244 T; 0 other;

Query Match 10.7%; Score 32; DB 22; Length 720;
Best Local Similarity 53.1%; Pred. No. 0.87;
Matches 68; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

OY 57 AATACATCATCATTTACCTGCGACTGTGGGGCTATGCGACTTAACTTTCGTCGCC 116
DB 594 AATCCCATCTCCAGAAATTTGATTTGGTCTGCTGAGAGAAATTAAGTAATTTGTC 535
OY 117 CCCGACGCTATGCAATAGACCAATTAAGTCAAAAGTCCGCTGATAGGCTGMAA 176
DB 534 TGTGAGTTTAAATGAATTTGAATTTAAATTTGAAGGAACATATGCTTTATTTTGAA 475
OY 177 AGTTCATT 184
DB 474 AGGACATT 467

RESULT 8

AA584698
ID AA584698 standard; cDNA: 765 BP.

XX AA584698;

DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #20502.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI: 2001-639362/73.

XX P-PSDB: ABG20511.

XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity

XX Claim 1; SEQ ID NO 20502; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving

identified in one or more metabolic pathways that lead to dwarfism and stunting in plants. Also included are vectors comprising the polynucleotides, transgenic plants (including the seed and leaf) transfected with the polynucleotides or vectors, a process for altering the metabolism of a plant comprising providing the above vector and a plant, and transfecting the plant with the vector under conditions such that the metabolism of the plant is altered by expression of the isolated nucleic acid from the vector, e.g. such that a stunting phenotype in an industrial plant is produced and a process for the characterisation of fractionated biological samples, comprising (a) providing one or more apparatus, a mass spectroscopy apparatus or data analysis software and (b) treating the fractionated biological samples and the reference samples with the gas chromatography apparatus to generate chromatographic data corresponding to the fractionated biological samples and the reference samples, (c) treating the mass spectroscopy apparatus to generate spectroscopic data corresponding to the fractionated biological samples and the reference samples and (d) processing the chromatographic and the spectroscopic data with the data analysis software. The nucleic acid and the vector are useful for altering the metabolism of a plant and for stunting a plant. The nucleic acids are useful in agriculture to create dwarf varieties of any plant species. The present sequence is a plant EST, expressed sequence tag, isolated from a cDNA library and related to stunting/dwarfism.

Sequence 1252 BP; 322 A; 267 C; 316 G; 347 T; 0 other;

Query Match 10.2%; Score 30.6; DB 24; Length 1252;

Best Local Similarity 50.3%; Pred. No. 3.5; Mismatches 75; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

29 TGATGATTATAGCAGTCCCTGCTGAAATATACATCATTCATTCGCGCTGCGCG 88
 1103 TTACGAGTTTCTCTTCAATCATCAAAATAGTTTCAGCTTTCATCTTGGGG 1162
 89 CTATCGACTTATGAGTTGCTGCGCCCTGAGTCTATGAGTATGAGCACTAACTGCA 148
 1163 TTTCGGAATTAATGAGTTTCTTCTTATGATGAAAGACTATGAGCACTTAATTA 1222
 149 AAAAAAGTCGCTGATAGGCTTGA 177
 1223 GTAGCTTCTCTTAATAAAAAA 1251

RESULT 11
 AAD00097 standard; cDNA; 1584 BP.

AC AAD00097;
 XX 31-JUL-2000 (first entry)
 XX Murine zsig49 cDNA.
 DE Murine zsig49 cDNA.
 XX Murine zsig49 cDNA; metabolic disease; diabetes; pancreatic disease;
 KW reproductive development; testicular function; wound healing;
 KW anti-microbial; antidiabetic; gene therapy; ss.
 XX Mus musculus.
 OS
 XX
 XX Key 1.1386 Location/Qualifiers
 FH CDS
 FT 1.1386
 FT sig_peptide
 FT 1.81
 FT /product= "Murine zsig49 protein"
 FT /tag= a
 FT 82..1383
 FT mat_peptide
 FT /tag= c
 FT /product= "Mature murine zsig49 protein"
 XX
 XX WO200023591-A2.

PD 27-APR-2000.
 XX
 XX 20-OCT-1999; 99MO-US24579.
 PF
 XX 21-OCT-1998; 98US-0176545.
 PR
 XX (ZYMO) ZYMOGENETICS INC.
 PA
 XX Sheppard PO, Holly RD, Gao Z, Whitmore TE, Maurer MF;
 PI WPI: 2000-339690/29.
 DR P-PSDB: AAY70815.
 DR
 XX zsig49 polypeptides and polynucleotides, useful for treatment and
 PT diagnosis of metabolic diseases such as diabetes -
 PS Claim 19; Page 126-128; 133pp; English.

The present cDNA sequence encodes murine zsig49 protein. zsig49 genes and proteins are useful for treatment and diagnosis of metabolic diseases such as diabetes, pancreatic conditions. They may exert regulatory effects on male gametes, reproductive development and testicular functions. They are therapeutically useful for aiding digestion, wound healing and anti-microbial applications.

Sequence 1584 BP; 467 A; 338 C; 374 G; 405 T; 0 other;

Query Match 10.2%; Score 30.6; DB 21; Length 1584;

Best Local Similarity 56.4%; Pred. No. 4; Mismatches 57; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 146 GCAAAAAAGTCCGCTGATAGGCTTGAAGAGTTCATTCACAGCCATTTCATCG 205
 422 GCAGAGACCTTCAATTCATTCGCGAGATATGATGATTCGAGACCACTTGTGAC 481
 QY 206 TACCCGATGAGGAGCGGCTGATGAGGCTTCTGCTGCTGCTGCTGCTGCTG 246
 482 TTGATGATGAGCGGCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 522

RESULT 12

AAC25839/c
 ID AAC25839 standard; cDNA; 443 BP.

AC AAC25839;
 XX 06-OCT-2000 (first entry)
 XX Human secreted protein 5' EST, SEQ ID NO: 29914.
 DE Human secreted protein 5' EST, SEQ ID NO: 29914.
 XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KW gene therapy; chromosome mapping; ss.
 XX Homo sapiens.
 OS
 XX EP1033401-A2.
 PN 06-SEP-2000.
 XX 21-FEB-2000; 2000EP-0200610.
 XX 26-FEB-1999; 99US-0122487.
 XX (GEST) GENSET.
 XX Dumas Milne Edwards J, Duclert A, Giordano J;
 XX WPI: 2000-500381/45.

New nucleic acid that is a 5' expressed sequence tag (5' EST) for obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for diagnostic, forensic, gene therapy and chromosome mapping procedures -

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

PR	19-MAY-2000	2000US-0205511
PR	07-JUN-2000	2000US-0209467
PR	28-JUN-2000	2000US-0219466
PR	30-JUN-2000	2000US-0215637
PR	07-JUL-2000	2000US-0216145
PR	11-JUL-2000	2000US-0216880
PR	11-JUL-2000	2000US-0217487
PR	14-JUL-2000	2000US-0217956
PR	26-JUL-2000	2000US-0218290
PR	26-JUL-2000	2000US-0220964
PR	14-AUG-2000	2000US-0224518
PR	14-AUG-2000	2000US-0224519
PR	14-AUG-2000	2000US-0225213
PR	14-AUG-2000	2000US-0225214
PR	14-AUG-2000	2000US-0225266
PR	14-AUG-2000	2000US-0225267
PR	14-AUG-2000	2000US-0225268
PR	14-AUG-2000	2000US-0225270
PR	14-AUG-2000	2000US-0225447
PR	14-AUG-2000	2000US-0225757
PR	14-AUG-2000	2000US-0225758
PR	18-AUG-2000	2000US-0225759
PR	22-AUG-2000	2000US-0226279
PR	22-AUG-2000	2000US-0226681
PR	22-AUG-2000	2000US-0226856
PR	23-AUG-2000	2000US-0227182
PR	30-AUG-2000	2000US-0227099
PR	01-SEP-2000	2000US-0228287
PR	01-SEP-2000	2000US-0228288
PR	01-SEP-2000	2000US-0228343
PR	01-SEP-2000	2000US-0229344
PR	05-SEP-2000	2000US-0229345
PR	05-SEP-2000	2000US-0229509
PR	06-SEP-2000	2000US-0229513
PR	06-SEP-2000	2000US-0230437
PR	08-SEP-2000	2000US-0230438
PR	08-SEP-2000	2000US-0231242
PR	08-SEP-2000	2000US-0231243
PR	08-SEP-2000	2000US-0231244
PR	08-SEP-2000	2000US-0231413
PR	08-SEP-2000	2000US-0231414
PR	08-SEP-2000	2000US-0232080
PR	12-SEP-2000	2000US-0232081
PR	12-SEP-2000	2000US-0231968
PR	14-SEP-2000	2000US-0232387
PR	14-SEP-2000	2000US-0232388
PR	14-SEP-2000	2000US-0232389
PR	14-SEP-2000	2000US-0232401
PR	14-SEP-2000	2000US-0233400
PR	14-SEP-2000	2000US-0233063
PR	14-SEP-2000	2000US-0233064
PR	14-SEP-2000	2000US-0233065
PR	21-SEP-2000	2000US-0234223
PR	21-SEP-2000	2000US-0234274
PR	25-SEP-2000	2000US-0234397
PR	25-SEP-2000	2000US-0234398
PR	25-SEP-2000	2000US-0234399
PR	26-SEP-2000	2000US-0235484
PR	27-SEP-2000	2000US-0235854
PR	27-SEP-2000	2000US-0235856
PR	29-SEP-2000	2000US-0236327
PR	29-SEP-2000	2000US-0236367
PR	29-SEP-2000	2000US-0236368
PR	29-SEP-2000	2000US-0236569
PR	02-OCT-2000	2000US-0236570
PR	02-OCT-2000	2000US-0236802
PR	02-OCT-2000	2000US-0237037
PR	02-OCT-2000	2000US-0237038
PR	02-OCT-2000	2000US-0237040
PR	13-OCT-2000	2000US-0239335
PR	13-OCT-2000	2000US-0239337
PR	20-OCT-2000	2000US-0240960

PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-02464617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250191.
 PR 05-DEC-2000; 2000US-0250193.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 (HUMA-) HUMAN GENOME SCI INC.
 PA Rosen CA, Barash SC, Ruben SM;
 XX
 XX
 PI
 DR WPI: 2001-457723/49.
 DR P-PDB: AAU18642.
 XX
 XX
 PT Isolated polypeptide for treating, preventing and/or prognosing
 PT respiratory disorders related to the lung including lung cancers and
 PT also for testing and detection e.g. diagnosis -
 XX
 PS Claim 1; SEQ ID No 101; 507pp: English.
 CC Sequences MAS29939-AS29930 represent CDNA molecules, which encode the
 CC lung antigen polypeptides of the invention. Lung antigen polypeptides and
 CC their associated polynucleotides are useful in the diagnosis, treatment

CC and prevention of various types of disorders in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. A pathological
CC condition can be determined by detecting the presence or absence of a
CC mutation in a lung antigen polynucleotide. The treatable disorders
CC include autoimmune diseases such as rheumatoid arthritis,
CC hyperproliferative disorders such as neoplasms of the breast or liver,
CC cardiovascular disorders such as cardiac arrest, cerebrovascular
CC disorders such as cerebral ischaemia, nervous system disorders such as
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi,
CC ocular disorders such as corneal infection, endocrine disorders such as
CC premature labour and infertility, gastrointestinal disorders such as
CC Crohn's disease, renal disorders such as glomerulonephritis and
CC respiratory disorders such as asthma and pleurisy. The polypeptides can
CC also be used to aid wound healing, to prevent skin aging due to sunburn,
CC to maintain organs before transplantation, to regenerate tissues and in
CC chemotaxis. The polypeptides can also be used as a food additive or
CC preservative to increase or decrease storage capabilities.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO

Query Match	Similarity	10.1%	Score 30.4	DB 22	Length 568
Best Local	Similarity	54.5%	Pred. No. 2.9		
Matches	61	Conservative	0	Mismatches	51
				Indels	0
				Gaps	0
QY	51	GCTGAAATAACATCATCATTCAGTCGCGACCTGGCGGCGCTATCGCACCTTAACGTTCCG	110		
DB	255	GTTAAAGACTAGACGCGCTGGATTATTTGCTTCTTCGCGGTAAGAAACATTTGCTTCC	196		
QY	111	GCTGCCCCCTCAGTCTATGCAATAGACATTAACGTCAAAAAAATGCGCT	162		
DB	195	CTGCCCCCAAGTCTTGAATGACACCTTTCTTTCACCAAAAAAGACACT	144		
RESULT 14					
ABA21158					
ID	ABA21158	standard	DNA	1579	BP.
XX	ABA21158				
AC					
XX					
DT	23-JAN-2002	(first entry)			
XX					
DE	Human nervous system related polynucleotide SEQ ID NO 13489.				
XX					
XX	Human: nootropic; neuroprotective; cytosolic; dermatological; virucide;				
KW	immunosuppressive; antinflammatory; anti-HIV; antibacterial; vulnery;				
KW	antiparkinsonian; antischling; antianemic; antiarthritic; cancer;				
KW	antihemematic; hepatoprotective; cerebroprotective; antiinflammatory;				
KW	antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;				
KW	antiparasitic; cardiact; immune disorder; cardiovascular disorder;				
KW	neurological disease; infection; nephrotoxic; gene therapy; vaccine; ds.				
XX					
OS	Homo sapiens.				
XX					
PN	WO200159063-A2.				
XX					
PD	16-AUG-2001.				
XX					
PF	17-JAN-2001; 2001WO-US01334.				
XX					
PR	31-JAN-2000; 2000US-0179065.				
PR	04-FEB-2000; 2000US-0180628.				
PR	24-FEB-2000; 2000US-0184664.				
PR	02-MAR-2000; 2000US-0186350.				
PR	16-MAR-2000; 2000US-0189874.				
PR	17-MAR-2000; 2000US-0190076.				
PR	18-APR-2000; 2000US-0198123.				
PR	19-MAY-2000; 2000US-0205515.				
PR	07-JUN-2000; 2000US-0209467.				
PR	28-JUN-2000; 2000US-0214886.				
PR	30-JUN-2000; 2000US-0215135.				
PR	07-JUL-2000; 2000US-0216647.				
PR	07-JUL-2000; 2000US-0216880.				
PR	11-JUL-2000; 2000US-0217487.				

PR 11-JUL-2000; 2000US-0217496.
 PR 14-JUL-2000; 2000US-0218290.
 PR 26-JUL-2000; 2000US-0220963.
 PR 14-AUG-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225266.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-0225758.
 PR 18-AUG-2000; 2000US-0225759.
 PR 22-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226868.
 PR 23-AUG-2000; 2000US-0227182.
 PR 30-AUG-2000; 2000US-0227009.
 PR 01-SEP-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 05-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 06-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231411.
 PR 08-SEP-2000; 2000US-0231412.
 PR 08-SEP-2000; 2000US-0232080.
 PR 08-SEP-2000; 2000US-0232081.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 21-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239335.
 PR 13-OCT-2000; 2000US-0239337.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 20-OCT-2000; 2000US-0242221.

PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250391.
 PR 01-DEC-2000; 2000US-0251160.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 (HUMA-) HUMAN GENOME SCI INC.
 Rosen CA, Barash SC, Ruben SM;
 WPI; 2001-541565/60.
 Nucleic acids encoding 3224 human nervous system antigen polypeptides,
 useful for preventing, diagnosing and/or treating nervous system
 cancers and metastases -
 Disclosure; SEQ ID NO 13489; 1701pp + Sequence Listing; English.
 The invention relates to novel genes (ABA11004-ABA21534) and proteins
 (ABBA14678-ABBA18001) useful for preventing, treating or ameliorating
 medical conditions e.g. by protein or gene therapy. The genes are
 isolated from a range of human tissues disclosed in the specification.
 The nucleic acids, proteins, antibodies and (ant)agonists are useful
 in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
 and ovarian cancer and other cancers of the adrenal gland, bone, bone
 marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
 (b) immune disorders e.g. Addison's disease, allergies, autoimmune
 haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
 disease, multiple sclerosis, rheumatoid arthritis and ulcerative

CC colitis; (c) cardiovascular disorders such as myocardial ischemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1579 BP; 410 A; 346 C; 353 G; 470 T; 0 other;
Query Match 10.1%; Score 30.4; DB 22; Length 1579;
Best Local Similarity 54.5%; Pred. No. 4.6;
Matches 61; Conservative 0; Mismatches 51; Indels 0; Gaps 0;
OY 51 GCTGAATATACATCATACATGACATGCGACATGCGGCTATGCGACTTACGTTCT 110
DB 990 GTTAAGACTAGACGCTGATTTCTTCTTCTGCTAGAAACATGTGCTTCT 1049
OY 111 GCTGCCCTCAGTCTATGACATAGACCATTAACGCAAAAAGTCGCT 162
DB 1050 CTTGCCCCCAAGTCTTGAATGACACTTTCTTCCACCAAAAAGACGCT 1101
RESULT 15
AAS30161/C
ID AAS30161 standard; DNA; 1579 BP.
XX
AC AAS30161;
XX
DT 21-NOV-2001 (first entry)
XX
DE Human lung antigen genomic DNA #231.
XX
KW Lung antigen protein; human; mouse; rabbit; goat; horse; cat; dog;
KW chicken; sheep; immunosuppressive; antiarthritic; vasotropic;
KW antitumoric; antiproliferative; cytostatic; cardiant; neuroprotective;
KW cerebroprotective; neurotrophic; antibacterial; virucide; fungicide; cancer;
KW ophthalmological; vulnary; gene therapy; autoimmune disease; neoplasm;
KW hyperproliferative disorder; breast; liver; cardiovascular disorder; ds;
KW cerebrovascular disorder; nervous system disorder; bacterial infection;
KW fungal infection; viral infection; ocular disorder; endocrine disorder;
KW gastrointestinal disorder; renal disorder; respiratory disorder;
KW wound healing; skin aging; organ transplantation; food preservative;
KW tissue regeneration; anti-fertility; food additive.
XX
OS Homo sapiens.
XX
PN WO200155303-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001MO-US01301.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.

PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 22-AUG-2000; 2000US-0227009.
PR 23-AUG-2000; 2000US-0228924.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0228287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 25-SEP-2000; 2000US-0235484.
PR 26-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 29-SEP-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 13-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241121.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 5, 2003, 17:51:17 ; Search time 54.3956 Seconds
(without alignments)
1691.369 Million cell updates/sec

Title: US-09-990-099-21

Perfect score: 300
Sequence: 1 atcatctgtcgtacgtcgtg99.....gattccagagcattcatt 300

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 441362 segs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 1008
Listing first 45 summaries

Database:

Issued Patents NA: *
1: /cgn2_6/prodata/1/lna/5A.COMB.seq:*
2: /cgn2_6/prodata/1/lna/5B.COMB.seq:*
3: /cgn2_6/prodata/1/lna/6A.COMB.seq:*
4: /cgn2_6/prodata/1/lna/6B.COMB.seq:*
5: /cgn2_6/prodata/1/lna/PCTUS.COMB.seq:*
6: /cgn2_6/prodata/1/lna/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match length	ID	Description
1	29.8	9.9	2437 1 US-08-456-647B-3	Sequence 3, Appl
2	29.8	9.9	2437 2 US-08-237-401A-3	Sequence 3, Appl
3	29.4	9.8	481 1 US-08-480-784-16	Sequence 16, Appl
4	29.4	9.8	481 1 US-08-483-553-16	Sequence 16, Appl
5	29.4	9.8	481 1 US-08-487-002-16	Sequence 16, Appl
6	29.4	9.8	481 1 US-08-483-554B-16	Sequence 16, Appl
7	29.4	9.8	481 1 US-08-488-011B-16	Sequence 16, Appl
8	29.4	9.8	481 4 US-08-850-727-16	Sequence 16, Appl
9	29.4	9.8	481 5 PCT-US95-10202-16	Sequence 16, Appl
10	29.4	9.8	481 5 PCT-US95-10203-16	Sequence 16, Appl
11	29.4	9.8	481 5 US-09-023-731-2	Sequence 2, Appl
12	29.4	9.8	1701 4 US-09-060-756-612	Sequence 612, App
13	28.2	9.7	349 4 US-09-194-613-4	Sequence 4, Appl
14	28.6	9.5	6534 4 US-09-177-461-1	Sequence 1, Appl
15	28.6	9.5	43804 4 US-08-294-312B-3	Sequence 3, Appl
16	28.4	9.3	3063 4 US-08-468-024B-3	Sequence 3, Appl
17	28	9.3	3063 4 US-09-641-638-81	Sequence 81, Appl
18	27.8	9.3	741 4 US-08-948-705-3	Sequence 3, Appl
19	27.4	9.1	1298 3 US-07-691-079C-8	Sequence 18, Appl
20	27.2	9.1	231 1 US-08-433-037-18	Sequence 27, Appl
21	27.2	9.1	1056 4 US-09-364-230-27	Sequence 14, Appl
22	27.2	9.1	1056 4 US-08-961-527-14	Sequence 14, Appl
23	26.8	8.9	2520 4 US-08-448-603A-31	Sequence 31, Appl
24	26.8	8.9	2570 4 US-09-134-075-31	Sequence 31, Appl
25	26.8	8.9	2570 4 US-09-492-739-31	Sequence 31, Appl
26	26.8	8.9	2570 4 US-08-448-603A-29	Sequence 29, Appl
27	26.8	8.9	2573 2	Sequence 29, Appl

28	26.8	8.9	2573 3 US-09-134-075-29	Sequence 29, Appl
29	26.8	8.9	2573 4 US-09-492-739-29	Sequence 29, Appl
30	26.8	8.9	3671 4 US-08-932-787B-5	Sequence 5, Appl
31	26.8	8.9	3671 4 US-08-932-012C-5	Sequence 5, Appl
32	26.8	8.9	3671 4 US-08-888-818C-5	Sequence 5, Appl
33	26.8	8.9	3767 4 US-08-928-941D-28	Sequence 28, Appl
34	26.8	8.9	3767 4 US-08-928-941D-30	Sequence 30, Appl
35	26.8	8.9	3767 4 US-09-280-590A-28	Sequence 28, Appl
36	26.8	8.9	3767 4 US-09-280-590A-30	Sequence 30, Appl
37	26.6	8.9	4602 3 US-08-904-452-3	Sequence 3, Appl
38	26.6	8.9	4602 4 US-09-517-639-3	Sequence 3, Appl
39	26.4	8.8	949 4 US-08-221-017B-615	Sequence 615, App
40	26.4	8.8	2137 3 US-08-821-818-4	Sequence 4, Appl
41	26.4	8.8	24417 2 US-08-846-762-1	Sequence 1, Appl
42	26.2	8.7	1668 1 US-08-463-090B-1	Sequence 1, Appl
43	26.2	8.7	13977 4 US-09-484-970B-60	Sequence 60, Appl
44	26.2	8.7	4403765 4 US-09-103-840A-2	Sequence 2, Appl
45	26.2	8.7	4411529 4 US-09-103-840A-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-456-647B-3
Sequence 3, Application US/08456647B
Patent No. 5811516
GENERAL INFORMATION:
APPLICANT: Lemke Ph.D. et al., Greg E.
TITLE OF INVENTION: PROTEIN-TYROSINE KINASE GENES
NUMBER OF SEQUENCES: 54
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 4225 Executive Square, Suite 1400
CITY: La Jolla
STATE: CA
COUNTRY: US
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/456,647B
FILING DATE: 02-JUN-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/237,401
FILING DATE: 02-MAY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/884,486
FILING DATE: 15-MAY-1992
ATTORNEY/AGENT INFORMATION:
NAME: Wetherell Ph.D., John R.
REGISTRATION NUMBER: 31,678
REFERENCE/DOCKET NUMBER: 07251/007002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 678-5099
TELEFAX: (619) 678-5070
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2437 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
IMMEDIATE SOURCE:
CLONE: Tyro-2
FEATURE:
NAME/KEY: CDS
LOCATION: 3..2118
US-08-456-647B-3

1617 CGAGGAGACTGGATGAAGAAGGCTACATGACTCCATGCATGCATGAC 1661

S-08-480-784-16

```
Query Match          9.8%; Score 29.4; DB 1; Length 481
Best Local Similarity 53.1%; Pred. No. 0.65;
```


LENGTH: 481 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
US-08-487-002-16

Query Match 9.8%; Score 29.4; DB 1; Length 481;
Best Local Similarity 53.1%; Pred. No. 0.65;
Matches 60; Conservative 1; Mismatches 52; Indels 0; Gaps 0;

QY 107 TCGTGTGCCCCCTCAGTCATGACATAGACCAATTAACAGTCAAAAAAGTCCGCTGATA 166
DB 122 TCCAGCTGGGACACAGAGTAGMCCCTGTCTCAAAAAAGTCAAAAAAGTCTGTG 63
QY 167 AGGCTGAAAGTTCATTTCAGACCCATTTCATTCGTCAGCCGATGAGAC 219
DB 62 AAAAATTAAAGTTCAATTATTCCTCACAATCTCTTTGACCATGTGCTC 10

RESULT 6
US-08-483-554B-16/C
Sequence 16, Application US/08483554B
Patent No. 5747282
GENERAL INFORMATION:

APPLICANT: Skolnick, Mark H.
APPLICANT: Goldgar, David E.
APPLICANT: Miki, Yoshio
APPLICANT: Swenson, Jeff
APPLICANT: Kamb, Alexander
APPLICANT: Harshman, Keith D.
APPLICANT: Shattuck-Eidens, Donna M.
APPLICANT: Tavligian, Sean V.
APPLICANT: Wiseman, Roger W.
APPLICANT: Futreal, P. Andrew
TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer
NUMBER OF SEQUENCES: 85
CORRESPONDENCE ADDRESS:
ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
STREET: 1201 New York Avenue, N.W., Suite 1000
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/483,554B
FILING DATE: 07-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/409,305
FILING DATE: 24-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/348,824
FILING DATE: 29-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/308,104
FILING DATE: 16-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/300,266
FILING DATE: 02-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/289,221
FILING DATE: 12-AUG-1994

ATTORNEY/AGENT INFORMATION:
NAME: Ihnen, Jeffrey L.
REGISTRATION NUMBER: 28,957
REFERENCE/DOCKET NUMBER: 24884-109347
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-962-4810
TELEFAX: 202-962-8300
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 481 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
US-08-483-554B-16

Query Match 9.8%; Score 29.4; DB 1; Length 481;
Best Local Similarity 53.1%; Pred. No. 0.65;
Matches 60; Conservative 1; Mismatches 52; Indels 0; Gaps 0;

QY 107 TCGTGTGCCCCCTCAGTCATGACATAGACCAATTAACAGTCAAAAAAGTCCGCTGATA 166
DB 122 TCCAGCTGGGACACAGAGTAGMCCCTGTCTCAAAAAAGTCAAAAAAGTCTGTG 63
QY 167 AGGCTGAAAGTTCATTTCAGACCCATTTCATTCGTCAGCCGATGAGAC 219
DB 62 AAAAATTAAAGTTCAATTATTCCTCACAATCTCTTTGACCATGTGCTC 10

RESULT 7
US-08-488-011B-16/C
Sequence 16, Application US/08488011B
Patent No. 5753441
GENERAL INFORMATION:

APPLICANT: Skolnick, Mark H.
APPLICANT: Goldgar, David E.
APPLICANT: Miki, Yoshio
APPLICANT: Swenson, Jeff
APPLICANT: Kamb, Alexander
APPLICANT: Harshman, Keith D.
APPLICANT: Shattuck-Eidens, Donna M.
APPLICANT: Tavligian, Sean V.
APPLICANT: Wiseman, Roger W.
APPLICANT: Futreal, P. Andrew
TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer
NUMBER OF SEQUENCES: 85
CORRESPONDENCE ADDRESS:
ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
STREET: 1201 New York Avenue, N.W., Suite 1000
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/488,011B
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/409,305
FILING DATE: 24-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/348,824
FILING DATE: 29-NOV-1994

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/308,104
FILING DATE: 16-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/300,266
FILING DATE: 02-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/289,221
FILING DATE: 12-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Ihnen, Jeffrey L.
REGISTRATION NUMBER: 28,957
REFERENCE/DOCKET NUMBER: 24884-109347-09
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-962-8300
TELEFAX: 202-962-8300
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 481 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHEICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
US-08-488-011B-16

Query Match 9.8%; Score 29.4; DB 1; Length 481;
Best Local Similarity 53.1%; Pred. No. 0.65;
Matches 60; Conservative 1; Mismatches 52; Indels 0; Gaps 0;
QY 107 TCGTGTGCCCCCTCATGCTATGCAATAGACCAATTAAGTCCGCTGATA 166
DB 122 TCCAGCTGGAGACAGAGTAAGMCCCTGTCTCAAAAAAAAAAAAGTACTGTG 63
QY 167 AGGCTGAAAGTTCATTTCCAGACCCATTTTACATCGTAGCCGATGAGAC 219
DB 62 AAAATTTAAGGTTCATTTATTTGCTCAGCATTCATCTCTTGACCATGTGTC 10

RESULT 8
US-08-850-727-16/c
Sequence 16, Application US/08850727
GENERAL INFORMATION:
APPLICANT: Skolnick, Mark H.
APPLICANT: Goldgar, David E.
APPLICANT: Miki, Yoshio
APPLICANT: Swenson, Jeff
APPLICANT: Kamb, Alexander
APPLICANT: Harshman, Keith D.
APPLICANT: Shattuck-Eidens, Donna M.
APPLICANT: Tavligian, Sean V.
APPLICANT: Wiseman, Roger W.
APPLICANT: Futreal, P. Andrew
TITLE OF INVENTION: 17q-linked Breast and Ovarian Cancer
TITLE OF INVENTION: Susceptibility Gene
NUMBER OF SEQUENCES: 85
CORRESPONDENCE ADDRESS:
ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
STREET: 1201 New York Avenue, N.W., Suite 1000
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/850,727
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/483,554
FILING DATE: 07-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/348,824
FILING DATE: 29-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/308,104
FILING DATE: 16-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/300,266
FILING DATE: 02-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/289,221
FILING DATE: 12-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Ihnen, Jeffrey L.
REGISTRATION NUMBER: 28,957
REFERENCE/DOCKET NUMBER: 24884-109347
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-962-8300
TELEFAX: 202-962-8300
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 481 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHEICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
US-08-850-727-16

Query Match 9.8%; Score 29.4; DB 4; Length 481;
Best Local Similarity 53.1%; Pred. No. 0.65;
Matches 60; Conservative 1; Mismatches 52; Indels 0; Gaps 0;
QY 107 TCGTGTGCCCCCTCATGCTATGCAATAGACCAATTAAGTCCGCTGATA 166
DB 122 TCCAGCTGGAGACAGAGTAAGMCCCTGTCTCAAAAAAAAAAAAGTACTGTG 63
QY 167 AGGCTGAAAGTTCATTTCCAGACCCATTTTACATCGTAGCCGATGAGAC 219
DB 62 AAAATTTAAGGTTCATTTATTTGCTCAGCATTCATCTCTTGACCATGTGTC 10

RESULT 9
PCT-US95-10202-16/c
Sequence 16, Application PC/TUS9510202
GENERAL INFORMATION:
APPLICANT: Shattuck-Eidens, Donna M.
APPLICANT: Simard, Jacques
APPLICANT: Eml, Mitsuru
APPLICANT: Nakamura, Yusuke
APPLICANT: Durocher, Francine
TITLE OF INVENTION: In Vivo Mutations and Polymorphisms
TITLE OF INVENTION: Susceptibility Gene
NUMBER OF SEQUENCES: 85
CORRESPONDENCE ADDRESS:
ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
STREET: 1201 New York Avenue, N.W., Suite 1000
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

```
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10202
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US
FILING DATE: 07-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/409,305
FILING DATE: 24-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/348,824
FILING DATE: 29-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08-308,104
FILING DATE: 16-SEP-1994
APPLICATION NUMBER: US 08/300,266
FILING DATE: 02-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/289,221
FILING DATE: 12-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Ihnen, Jeffrey L.
REGISTRATION NUMBER: 28,957
REFERENCE/DOCKET NUMBER: 24884-109347
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-962-4810
TELEFAX: 202-962-8300
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 481 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
PCT-US95-10202-16

Query Match
Best Local Similarity 9.8%; Score 29.4; DB 5; Length 481;
Matches 60; Conservative 1; Mismatches 52; Indels 0; Gaps 0;

QY 107 TCGTCTGCCCCCTAGTCTATGCAATAGACATAAAGTCAAAAAAAGTCCGCTGATA 166
      |||||  |||  |||  |||  |||  |||||  |||||  |||||
DB 122 TCCAGCTGGGACACAGAGTAGTCCCTGTCTCAAAAAAAGTAGAGTGTG 63
      |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY 167 AGGCTTGAAGTTCATTCCAGACCATTTTACATGCTGACCGATGAGAC 219
      |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 62 AAAATTTAAGTTCATTATTTGCTCACAATCTATCTTTGACCATGTGTC 10
      |||||  |||||  |||||  |||||  |||||  |||||  |||||

RESULT 10
PCT-US95-10203-16/c
; Sequence 16, Application PC/TUS9510203
; GENERAL INFORMATION:
; APPLICANT: Skolnick, Mark H.
; APPLICANT: Goldgar, David E.
; APPLICANT: Miki, Yoshio
; APPLICANT: Swenson, Jeff
; APPLICANT: Kamb, Alexander
; APPLICANT: Harshman, Keith D.
; APPLICANT: Shattuck-Eidens, Donna M.
; APPLICANT: Tavligian, Sean V.
; APPLICANT: Wiseman, Roger W.
; APPLICANT: Futreal, P. Andrew
; TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer
```

```
TITLE OF INVENTION: Susceptibility Gene
NUMBER OF SEQUENCES: 85
CORRESPONDENCE ADDRESS:
ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
STREET: 1201 New York Avenue, N.W., Suite 1000
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10203
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US
FILING DATE: 07-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/409,305
FILING DATE: 24-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/348,824
FILING DATE: 29-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08-308,104
FILING DATE: 16-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/300,266
FILING DATE: 02-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/289,221
FILING DATE: 12-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Ihnen, Jeffrey L.
REGISTRATION NUMBER: 28,957
REFERENCE/DOCKET NUMBER: 24884-109347
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-962-4810
TELEFAX: 202-962-8300
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 481 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
PCT-US95-10203-16

Query Match
Best Local Similarity 9.8%; Score 29.4; DB 5; Length 481;
Matches 60; Conservative 1; Mismatches 52; Indels 0; Gaps 0;

QY 107 TCGTCTGCCCCCTAGTCTATGCAATAGACATAAAGTCAAAAAAAGTCCGCTGATA 166
      |||||  |||  |||  |||  |||  |||||  |||||  |||||
DB 122 TCCAGCTGGGACACAGAGTAGTCCCTGTCTCAAAAAAAGTAGAGTGTG 63
      |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY 167 AGGCTTGAAGTTCATTCCAGACCATTTTACATGCTGACCGATGAGAC 219
      |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 62 AAAATTTAAGTTCATTATTTGCTCACAATCTATCTTTGACCATGTGTC 10
      |||||  |||||  |||||  |||||  |||||  |||||  |||||

RESULT 11
PCT-US95-10220-16/c
; Sequence 16, Application PC/TUS9510220
; GENERAL INFORMATION:
```

Query Match	9.88;	Score 29.4;	DB 5;	Length 481;
Best Local Similarity	53.13;	Pred. No. 0.65;		
Matches 60;	Conservative	1;	Mismatches 52;	Indels 0;
			Gaps	
107	TTCTGGCTGCCCTCACTGCTATGCAATGACCATTAACATGCAAAAAAGGCTCGATGA	166		
07	TTCTGGCTGCCCTCACTGCTATGCAATGACCATTAACATGCAAAAAAGGCTCGATGA	166		

```

1  APPLICANT: Billault, Alain
2  TITLE OF INVENTION: METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST FROM
3  TITLE OF INVENTION: THE GENOME OF A MYCOBACTERIUM USING A PAC-BASED DNA
4  TITLE OF INVENTION: LIBRARY APPLICATION TO THE DETECTION OF MYCOBACTERIA
5  FILE REFERENCE: 3495-0169
6  CURRENT APPLICATION NUMBER: US/09/060,756
7  NUMBER OF SEQ ID NOS: 743
8  SOFTWARE: PatentIn Ver. 2.0
9  SEQ ID NO 612
10 LENGTH: 349
11 TYPE: DNA
12 ORGANISM: Mycobacterium tuberculosis
13 FEATURE:
14 NAME/KEY: unsure
15 LOCATION: (various positions within the sequence)
16 OTHER INFORMATION: applicants are uncertain of bases designated as "n"
17 US-09-060-756-612

```

	Query Match	9.7%;	Score 29.2;	DB 4;	Length 349;
	Best Local Similarity	57.0%;	Pred. No. 0.65;		
	Matches	49;	Conservative	0;	Mismatches 37; Indels 0; Gaps 0
Qy	9 GTTATCGTGGGCTTTCGCCATGATTATAGCCATCCCTTGCAGAAATACATATC				
Db	287 GTNACACAGTATGTCGCCGCCCTTTTGTGAGATCTTCGACGCCGTGAGCAAGGTGACATCTTT				
Qy	69 ATTACGTCGCGACGTGCGCGCATATCG				
Db	227 GTTCGCTGCACATGTGAGCAGCGTCG				

```

1      RESULT 14
2      US-09-194-613-d/c
3      ; Sequence 4, Application US/09194613
4      ; Patent No. 6251654
5      GENERAL INFORMATION:
6      ;
7      ; APPLICANT: GORDON, Karl H.
8      ; APPLICANT: HANZLIK, Terry N.
9      ; TITLE OF INVENTION: MODIFIED SMALL RNA VIRUSES
10     ; NUMBER OF SEQUENCES: 22
11     ; CORRESPONDENCE ADDRESS:
12     ; ADDRESSEE: Mcdermott, Will & Emery
13     ; STREET: 600 13th Street, N.W.
14     ; CITY: Washington
15     ; STATE: D.C.
16     ; COUNTRY: USA
17     ; ZIP: 20005-3096
18     COMPUTER READABLE FORM:
19     ;
20     ; MEDIUM TYPE: Floppy disk
21     ; COMPUTER: IBM PC compatible
22     ; OPERATING SYSTEM: PC-DOS/MS-DOS
23     ; SOFTWARE: PatentIn Release #1.0, Version #1.30
24     ; CURRENT APPLICATION DATA:
25     ; APPLICATION NUMBER: US/09/194,613
26     ; FILING DATE: 30-NOV-1998
27     ; CLASSIFICATION: 424
28     ATTORNEY/AGENT INFORMATION:
29     ;
30     ; NAME: Bucca Ph.D., Daniel
31     ; REGISTRATION NUMBER: 42,368
32     ; REFERENCE/DOCKET NUMBER: 50179-061
33     ; TELECOMMUNICATION INFORMATION:
34     ; TELEPHONE: 202-756-8000
35     ; TELEFAX: 202-756-8087
36     ; INFORMATION FOR SEQ. ID NO.: 4:
37     ; SEQUENCE CHARACTERISTICS:
38     ; LENGTH: 6534 base pairs
39     ; TYPE: nucleic acid
40     ; STRANDEDNESS: double
41     ; TOPOLOGY: linear
42     ; MOLECULE TYPE: CDNA
43     ; US-09-194-613-4

```

	Query Match	Best Local Similarity	Score 28.6;	DB 4;	Length 6534;
Matches	55;	Conservative	0;	Mismatches	44;
				Indels	0;
QY	17	TGGGTGGGCGCCATGAGTTATAGCGATCCCTGCTGAAATATACATCATCATAGTC	76		
DB	1515	TGGGATATTTCCCTTTGGAGCTCGCGAATTCGGATGATGGCAGCAAAATGATTTCTTGGAC	1456		
QY	77	GCAATGTGGGGCGGATACGACACTTAAAGCTTTTGGTCTGC	115		
DB	1455	CCAACTCGCGACAGATGAGGTTATATAGCCTGTGGTGTGC	1417		

```

RESULT 15
US-09-171-461-1/c
: Sequence 1, Application US/09171461
: Patent No. 6335016
: GENERAL INFORMATION:
: APPLICANT: Baker, Adam
: APPLICANT: Cotten, Matthew
: APPLICANT: Chioocca, Susana
: APPLICANT: Kurzbaner, Robert
: APPLICANT: Schaffner, Gotthold
: TITLE OF INVENTION: Chicken Embryo Lethal Orphan (CELO) VIRUS
: FILE REFERENCE: 0652.1800000
: CURRENT APPLICATION NUMBER: US/09/171.461
: EARLIER FILING DATE: 1999-01-12
: EARLIER APPLICATION NUMBER: PCT/EP97/01944
: NUMBER OF SEQ ID NOS: 54
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1
: LENGTH: 43804
: TYPE: DNA
: ORGANISM: CELO VIRUS
: FEATURE:
: NAME/KEY: gene
: LOCATION: (12193)..(15043)
: OTHER INFORMATION: /gene: L1
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (15080)
: OTHER INFORMATION: /note= L2 region penton base splice acceptor site
: FEATURE:
: NAME/KEY: gene
: LOCATION: (15110)..(17495)
: OTHER INFORMATION: /gene: L2
: FEATURE:
: NAME/KEY: polyA_site
: LOCATION: (17526)
: FEATURE:
: NAME/KEY: gene
: LOCATION: (17559)..(21754)
: OTHER INFORMATION: /gene: L3
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (18261)
: OTHER INFORMATION: /gene: L3 /note= hexon splice acceptor site
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (21102)
: OTHER INFORMATION: /gene: L3 /note= protease splice acceptor site
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (21123)
: OTHER INFORMATION: /gene: L3 /note= protease splice acceptor site
: NAME/KEY: polyA_site
: LOCATION: (21767)
: FEATURE:
: NAME/KEY: polyA_site
: LOCATION: (21824)
: FEATURE:

```

```

NAME/KEY: polyA_site
LOCATION: (21836)
FEATURE:
NAME/KEY: polyA_site
LOCATION: (21882)
FEATURE:
NAME/KEY: misc_feature
LOCATION: (23608)
OTHER INFORMATION: /note= 100k splice acceptor site
FEATURE:
NAME/KEY: misc_feature
LOCATION: (23649)
OTHER INFORMATION: /note= 100k splice acceptor site
FEATURE:
NAME/KEY: gene
LOCATION: (23680)..(27886)
OTHER INFORMATION: /gene: L4
FEATURE:
NAME/KEY: polyA_site
LOCATION: (27920)
FEATURE:
NAME/KEY: misc_feature
LOCATION: (28315)
OTHER INFORMATION: /note= fibre splice acceptor site
FEATURE:
NAME/KEY: misc_feature
LOCATION: (28341)
OTHER INFORMATION: / note= fibre splice acceptor site
FEATURE:
NAME/KEY: gene
LOCATION: (28363)..(31768)
OTHER INFORMATION: /gene: L5
FEATURE:
NAME/KEY: misc_feature
LOCATION: (30511)
OTHER INFORMATION: /gene: L5 /note= fibre splice acceptor site
FEATURE:
NAME/KEY: polyA_site
LOCATION: (31770)
US-09-171-461-1

```

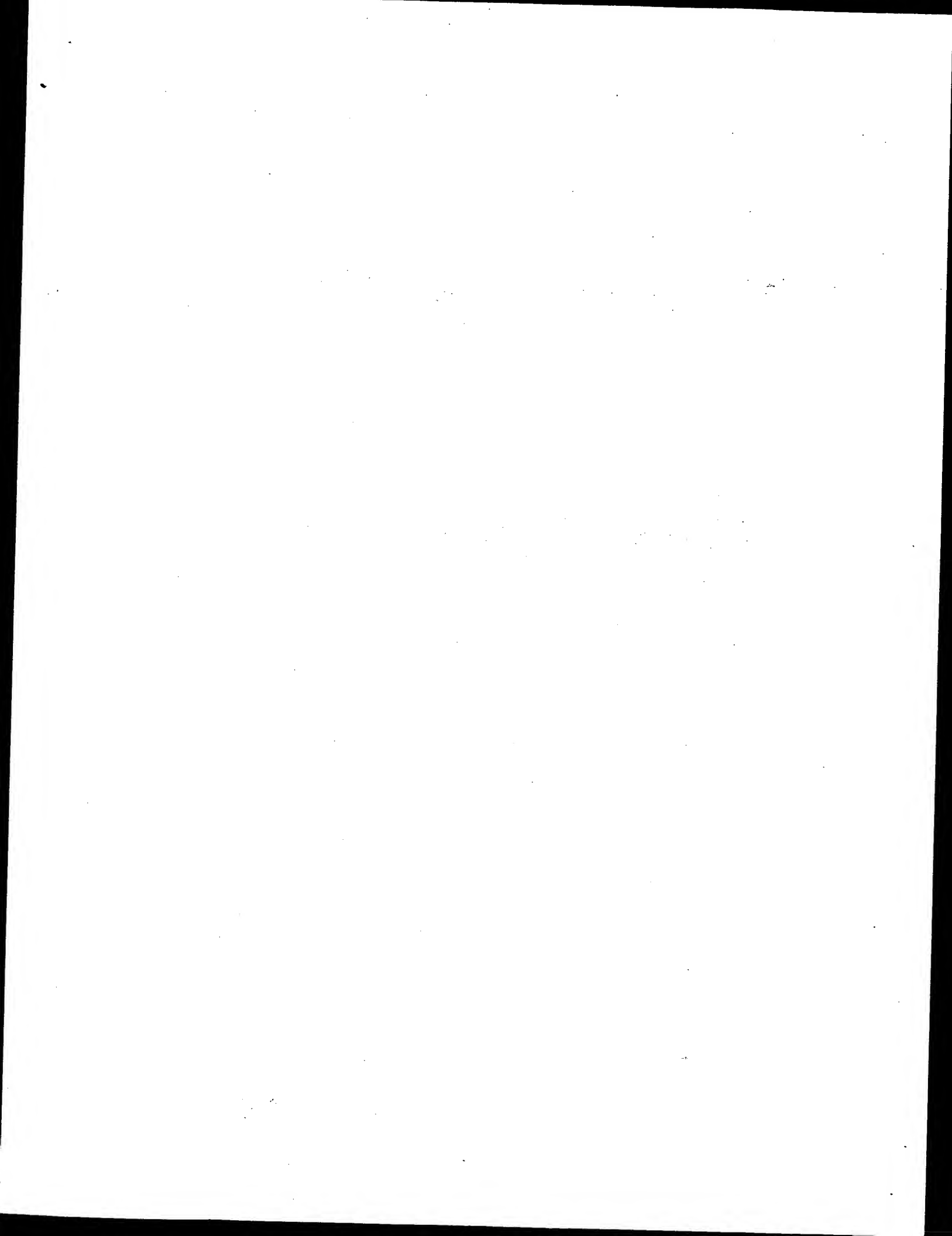
```

Query Match          9.5%; Score 28.4; DB 4; Length 43804;
Best Local Similarity 51.6%; Pred. No. 15;
Matches 65; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

OY 137 CCATTAAGTGCMAAAAAAGTCCCTGATAGGCTTGAAAGTTCATTCCAGACCCCATP 196
DB 41222 CCATTAATGCCCCATAGGGAATGATGTTCCCTTCAAGTACATTGCCGGAATGTA 41163
OY 197 TTACATCGTAGCGAGGAGCGGCGCTGATGGGTCTTCTGCTACCTGACCTGTCAT 256
DB 41162 CTTAATATGGAGCTTAGAGAGAGAGACAGAACAAAGTGTGTGGGTCCATCGCGCTCGC 41103
OY 257 TGTGGA 262
DB 41102 TGTGTA 41097

```

Search completed: May 5, 2003, 18:00:48
Job time : 63.3956 secs



Query Match	Best Local Similarity	12.0%;	Score 36;	DB 9;	Length 423;
Matches	72;	Conservative	0;	Mismatches	60;
				Indels	0;
				Gaps	0;
QY	153	AAAGTCGGCGATGAAGGCTTGAAGAGTCATTTCACAGCCCATTTTACATCGTACC	GA	212	
Db	5	AAAGCGTCGCGTTGGCGCTGAGAGATCCATCTATTATTTACCAACATTAAGACATGAAC	64		
QY	213	TGAGAGCGCCCTGCATCGGCTGTCTGGCTACCTGCACCTGCTCATGTGTGGAAGGCTTACA	272		
Db	65	TGTGGAGAGACCTGCCTGGATATCTCTCCAAACCAAGGTATGCTGCTCTCTATCA	124		
QY	273	TTCTCGCTGATT	284		
Db	125	ACCATTCCAATT	136		

RESULT 2
US-09-925-297-225 Application us/09925297
; Sequence 225
; Patent No. US20020081659A1

```

; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA105
; CURRENT APPLICATION NUMBER: US/09/925,297
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05989
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; NUMBER OF SEQ ID NOS: 928
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 225
; LENGTH: 1791
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-297-225

```

```

Query Match      11.5%; Score 34.4; DB 10; Length 1791;
Best Local Similarity 53.8%; Pred. No. 0.14;
Matches 71; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

```

```

QY 153 AAGTCGGCTGATGAAGCTTGAAGTTCATTCAGACCCATTTCATGCTAGCCCA 212
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 276 AAGGCTGCTGTTGGCTGAGAGCCATTATATTCACACATTAAAGACATGAGC 335
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 213 TGAGGACGGCCTGATGGGTGCTGCTACCTGACCTGCTTCATTTGAGGCTTTACA 272
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 336 TGTGGAAGACCCCTCGATATTCCTGCAACCCAGAGGTGCTGCTCTATCA 395
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 273 TTTCTGCTGATT 284
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 396 ACCATTCCAATT 407
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 3
US-10-041-856-1/C
; Sequence 1, Application US/10041856
; Patent No. US2002016929A1
; GENERAL INFORMATION:
; APPLICANT: SLAUGENHAUPT, SUSAN
; APPLICANT: GUSELLA, JAMES F.
; TITLE OF INVENTION: GENE FOR IDENTIFYING INDIVIDUALS WITH FAMILIAL
; FILE REFERENCE: 1829-4004US1
; CURRENT APPLICATION NUMBER: US/10/041,856
; PRIOR FILING DATE: 2002-07-08
; PRIOR APPLICATION NUMBER: 60/260,080
; NUMBER OF SEQ ID NOS: 88
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 66479
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-041-856-1

```

```

Query Match      11.3%; Score 34; DB 9; Length 66479;
Best Local Similarity 56.1%; Pred. No. 1.2;
Matches 64; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

```

```

QY 1 ATTCACTGTTGATCGGTGGTGGCTGATGATTATAGCGATCCCTGCTGAATAA 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 50954 ATTATCAATTTTATGATGTTGTTGATTAATTAATCAATCAATTCATTCGA 50895
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 61 ACATCATCATTAAGTGCATGCTGCGGCTATGCGACCTTAACGTTTCTGCTG 114
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 50894 AAGTCATGAAGATTAGCCCTGTTTCTTAAGAGTTTATAGTTGCTGCTG 50841
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 4
US-10-003-806-10/C
; Sequence 10, Application US/10003806

```

```

; Patent No. US20020119929A1
; GENERAL INFORMATION:
; APPLICANT: Bishop, Colin E.
; APPLICANT: Agoulrik, Alexander I.
; APPLICANT: Zhu, Qichao
; TITLE OF INVENTION: CAN 1 AND ITS ROLE IN MAMMALIAN INFERTILITY
; FILE REFERENCE: P02066US1/10024824
; CURRENT APPLICATION NUMBER: US/10/003,806
; PRIOR FILING DATE: 2001-11-02
; PRIOR APPLICATION NUMBER: 60/245,872
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 173808
; TYPE: DNA
; ORGANISM: Mouse
US-10-003-806-10

```

```

Query Match      10.8%; Score 32.4; DB 12; Length 173808;
Best Local Similarity 62.2%; Pred. No. 7.1;
Matches 51; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

```

```

QY 127 ATGCATATGACCATTAACGCAAAAAAGTCGCTGATAGGCTTGAAGATTCATTTC 186
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 116803 AGGAAAAAGATTAAACTACAAAAACAATCAGCCAAAAAGAGATGCAAAAGATGTTCC 116744
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 187 CAGACCAATTTTACATGCTAG 208
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 116743 CAGAACAAATGGTAAATCATCATAG 116722
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 5
US-09-910-664-114
; Sequence 114, Application US/09910664
; Publication No. US2002019464A1
; GENERAL INFORMATION:
; APPLICANT: POGUE, Greg P.
; APPLICANT: DELLA-CIOPPA, Guy R.
; APPLICANT: WOLFE, Gerson M.
; APPLICANT: ZHENG, Wenjin
; TITLE OF INVENTION: METHODS OF CREATING DMARF PHENOTYPES IN
; FILE REFERENCE: 00801018900US01
; CURRENT APPLICATION NUMBER: US/09/910,664
; PRIOR FILING DATE: 2001-07-20
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 114
; LENGTH: 1252
; TYPE: DNA
; ORGANISM: Nicotiana benthamiana
US-09-910-664-114

```

```

Query Match      10.2%; Score 30.6; DB 9; Length 1252;
Best Local Similarity 50.3%; Pred. No. 2.5;
Matches 75; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

```

```

QY 29 TCATAGATTATAGCATCCCTGCTGAATAATACATCATCATTCATGTCGCGCGG 88
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1103 TTACGAGTTTTCCTTCATCAATCAATCAATAGTTTCAAGCTTTCATTTCACTTTGGGG 1162
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 89 CTATGCGCATTTAAGGTTTGGTGGCTGCCCTCGAGTCTATGCAATAGACATAAAGTCA 148
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1163 TTTGGGATTTTAATGTTTCTTGTATGATGAAGACTATGCAATTAAGGACACTAATAAA 1222
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 149 AAAAAAAGTCCGCTGATAGAGCTTGAAAA 177
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1223 GTAAGCTTTCTTCTTAATAAAAAA 1251
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 6
US-09-923-876-2531

```


Query Match
Best Local Similarity 10.1%; Score 30.4; DB 9; Length 1579;
Matches 61; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

OY 51 GCTGAAATATACATCATTCATTCATTCGCGACCTGTGGCGCTATCCGACTTTAACTTTCGT 110
DB 590 GTTAAGACTAGCAGCGTGTGATTTGCTTCTGCTAGAAACATTTGTCTTCT 531
OY 111 GCTGCCCCCTCAGCTATGCAATAGCACTAACTGCAAAAAAAGTCCGCT 162
DB 530 CCTGCCCCCAAGTCTTGAATGACCTTTCTTTCACCAAAAAAGACAGCT 479

RESULT 10

US-09-764-878-425/c
; Sequence 425, Application US/09764878
; Patent No. US20020090615A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA121
; CURRENT APPLICATION NUMBER: US/09/764,878
; PRIOR FILING DATE: 2001-01-17
; NUMBER OF SEQ ID NOS: 428
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO: 425
; LENGTH: 1579
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-878-425

Query Match
Best Local Similarity 10.1%; Score 30.4; DB 10; Length 1579;
Matches 61; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

OY 51 GCTGAAATATACATCATTCATTCATTCGCGACCTGTGGCGCTATCCGACTTTAACTTTCGT 110
DB 590 GTTAAGACTAGCAGCGTGTGATTTGCTTCTGCTAGAAACATTTGTCTTCT 531
OY 111 GCTGCCCCCTCAGCTATGCAATAGCACTAACTGCAAAAAAAGTCCGCT 162
DB 530 CCTGCCCCCAAGTCTTGAATGACCTTTCTTTCACCAAAAAAGACAGCT 479

RESULT 11

US-10-079-854-422/c
; Sequence 422, Application US/10079854
; Publication No. US20030054368A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA121C1
; CURRENT APPLICATION NUMBER: US/10/079,854
; PRIOR FILING DATE: 2002-02-22
; NUMBER OF SEQ ID NOS: 428
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO: 422
; LENGTH: 4607
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-079-854-422

Query Match
Best Local Similarity 10.1%; Score 30.4; DB 9; Length 4607;
Matches 61; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

OY 51 GCTGAAATATACATCATTCATTCATTCGCGACCTGTGGCGCTATCCGACTTTAACTTTCGT 110
DB 590 GTTAAGACTAGCAGCGTGTGATTTGCTTCTGCTAGAAACATTTGTCTTCT 531

OY 111 GCTGCCCCCTCAGCTATGCAATAGCACTAACTGCAAAAAAAGTCCGCT 162
DB 530 CCTGCCCCCAAGTCTTGAATGACCTTTCTTTCACCAAAAAAGACAGCT 479

RESULT 12

US-10-079-854-424/c
; Sequence 424, Application US/10079854
; Publication No. US20030054368A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA121C1
; CURRENT APPLICATION NUMBER: US/10/079,854
; PRIOR FILING DATE: 2002-02-22
; NUMBER OF SEQ ID NOS: 428
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO: 424
; LENGTH: 4607
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-079-854-424

Query Match
Best Local Similarity 10.1%; Score 30.4; DB 9; Length 4607;
Matches 61; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

OY 51 GCTGAAATATACATCATTCATTCATTCGCGACCTGTGGCGCTATCCGACTTTAACTTTCGT 110
DB 590 GTTAAGACTAGCAGCGTGTGATTTGCTTCTGCTAGAAACATTTGTCTTCT 531
OY 111 GCTGCCCCCTCAGCTATGCAATAGCACTAACTGCAAAAAAAGTCCGCT 162
DB 530 CCTGCCCCCAAGTCTTGAATGACCTTTCTTTCACCAAAAAAGACAGCT 479

RESULT 13

US-09-764-878-422/c
; Sequence 422, Application US/09764878
; Patent No. US20020090615A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA121
; CURRENT APPLICATION NUMBER: US/09/764,878
; PRIOR FILING DATE: 2001-01-17
; NUMBER OF SEQ ID NOS: 428
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO: 422
; LENGTH: 4607
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-878-422

Query Match
Best Local Similarity 10.1%; Score 30.4; DB 10; Length 4607;
Matches 61; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

OY 51 GCTGAAATATACATCATTCATTCATTCGCGACCTGTGGCGCTATCCGACTTTAACTTTCGT 110
DB 590 GTTAAGACTAGCAGCGTGTGATTTGCTTCTGCTAGAAACATTTGTCTTCT 531
OY 111 GCTGCCCCCTCAGCTATGCAATAGCACTAACTGCAAAAAAAGTCCGCT 162
DB 530 CCTGCCCCCAAGTCTTGAATGACCTTTCTTTCACCAAAAAAGACAGCT 479

RESULT 14

US-09-764-878-424/c
; Sequence 424, Application US/09764878
; Patent No. US20020090615A1

```
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA121
; CURRENT APPLICATION NUMBER: US/09/764,878
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 428
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 424
; LENGTH: 4607
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-764-878-424

Query Match          10.1%; Score 30.4; DB 10; Length 4607;
Best Local Similarity 54.5%; Pred. No. 5.7;
Matches 61; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

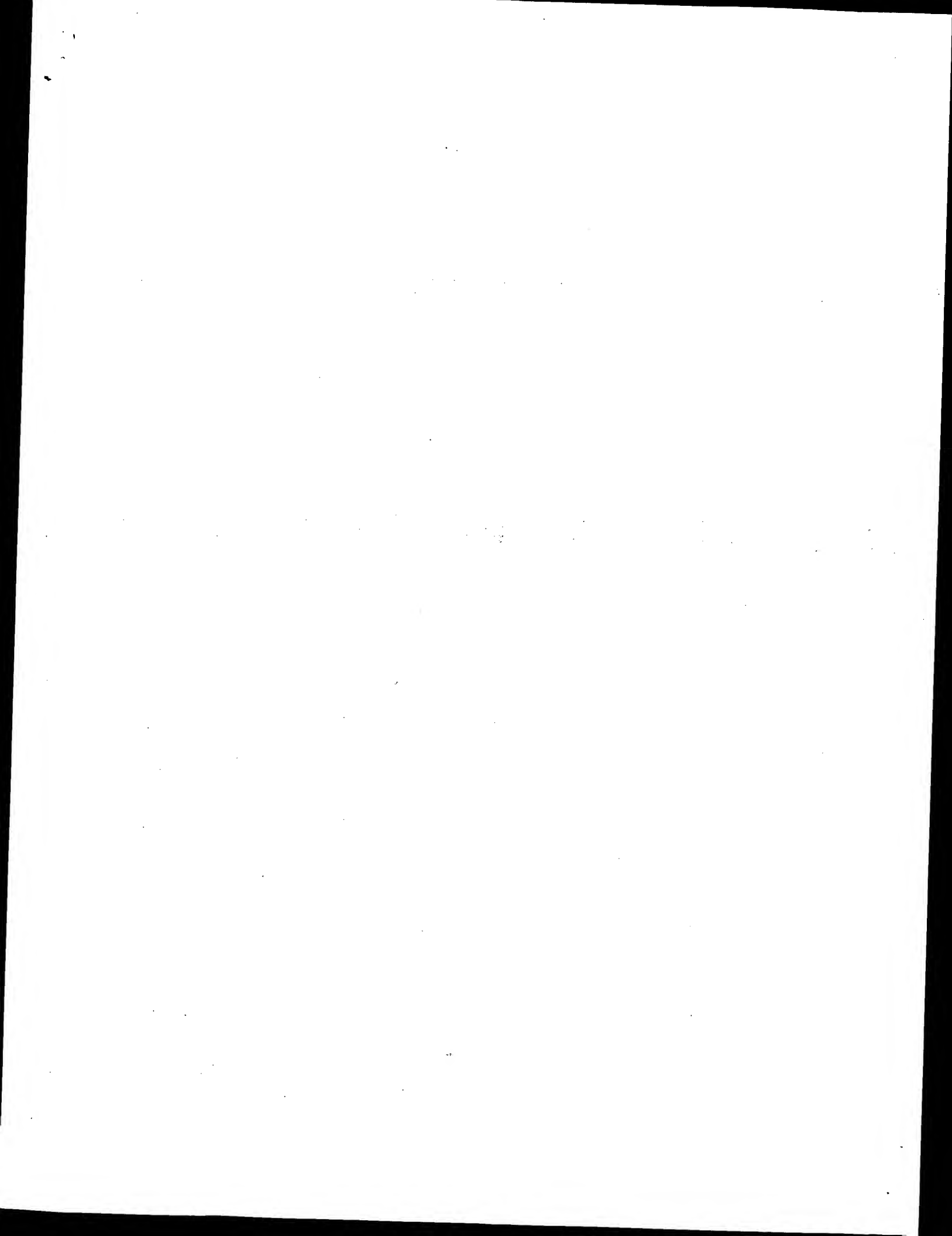
QY 51 GCTGAATAATACATCATTCATTCAGTCGCCACTGTGGCGCTATCGCACTTTAAAGTTTGGT 110
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 590 GTTAAGACTAGACGCGTGGATTATTTGTTCTTCTGCTAGAGAACATGTGCTTCTCT 531
QY 111 GCTGCCCCCTCAGCTATGCAATAGACCATTAACGCAAAAAAAGTCGCT 162
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 530 CTTGCCCCCAAGTTCTTGAATGCACCTTTCTTCACCAAAAAAGACGCT 479

RESULT 15
US-09-901-136-3
; Sequence 3, Application US/09901136
; Publication No. US20030039968A1
; GENERAL INFORMATION:
; APPLICANT: HU, Song et al.
; TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS, AND
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL001273
; CURRENT APPLICATION NUMBER: US/09/901,136
; CURRENT FILING DATE: 2001-07-10
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 378361
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)..(378361)
; OTHER INFORMATION: n - A,T,C or G
; US-09-901-136-3

Query Match          10.1%; Score 30.4; DB 9; Length 378361;
Best Local Similarity 57.3%; Pred. No. 53;
Matches 55; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 191 CCCATTTTACATGTAGCCGATGAGAGCGGCCCTGATGGGTGCTTACCTGACCT 250
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 188949 CTCACCTCAGACCTGTACCTTTTGTACTAGATAATCTTTGTTGTGGGCTGTGCT 189008
QY 251 GTCCATTGTGAGAGGCTTTACATTCGCTGATTTTC 286
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 189009 GTGCAATTGAGAGTGTAGCAGCATCCCTGATCTC 189044

Search completed: May 5, 2003, 18:05:29
Job time : 245.637 secs
```



GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 5, 2003, 17:53:33 ; Search time 1830.49 Seconds
(without alignments)
2654.281 Million cell updates/sec

Title: US-09-990-099-21

Perfect score: 300
Sequence: 1 attcatctgtgtatgctggtg.....gatttcagagcatatgatt 300

Scoring table: IDENTITY_NUC
Gapop 10.0 ; Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database:

EST:*
1: em_estba:*
2: em_esthm:*
3: em_estln:*
4: em_estlu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vit:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	length	DB ID	Description
C 1	131.4	43.8	965	17	AF035988
C 2	92.6	30.9	863	17	AF094932
C 3	36.4	12.1	313	10	BB791480
C 4	36	12.0	690	10	BE298245
C 5	35.6	11.9	968	17	CNS03VR3
C 6	35.2	11.7	829	13	BI763074

7	34.4	11.5	373	9	AA336439
8	34.4	11.5	521	12	BF727196
9	34.4	11.5	557	10	BE386483
10	34.4	11.5	623	13	BI869063
11	34.4	11.5	728	9	AL529169
12	34.4	11.5	769	12	BG680800
13	34.4	11.5	769	12	BI764142
14	34.4	11.5	827	13	BI819186
15	34.4	11.5	866	9	AL117405
16	34.4	11.5	872	9	AL530997
17	34.4	11.5	900	14	BQ232193
18	34.4	11.5	952	13	BI517912
19	34.4	11.5	983	14	BQ868892
20	34.4	11.5	1038	9	AL545606
21	34.4	11.5	1063	14	BM905328
22	34.4	11.5	1105	14	BM928240
23	34.4	11.5	1117	14	BM924731
24	34.4	11.5	1133	14	BM924731
25	34.2	11.4	605	13	BM492118
26	34	11.3	666	17	AZ834122
27	33.8	11.3	351	14	BQ283149
28	33.8	11.3	440	9	AI285962
29	33.8	11.3	453	10	BE363625
30	33.8	11.3	595	10	BE595980
31	33.6	11.2	1084	17	CNS05909
32	33.4	11.1	630	10	AW955927
33	33	11.0	572	17	AO670967
34	32.8	10.9	381	13	BI416919
35	32.8	10.9	387	12	BF900486
36	32.8	10.9	502	10	AM602876
37	32.8	10.9	515	13	BM212354
38	32.8	10.9	525	10	AV915911
39	32.8	10.9	585	14	BQ291456
40	32.8	10.9	666	10	AV915809
41	32.8	10.9	681	10	BE438175
42	32.8	10.9	1011	17	CNS04CVP
43	32.6	10.9	1050	17	CNS020USI
44	32.6	10.8	511	17	AZ873575
45	32.4	10.8	575	14	BQ634660

ALIGNMENTS

RESULT 1
AF035988/c 965 bp DNA linear GSS 29-AUG-2000
LOCUS AF035988 Salmonella typhimurium LT2, Lambda DASH II Salmonella
DEFINITION typhimurium genomic clone 7-T7, DNA sequence.

ACCESSION AF035988
VERSION AF035988
KEYWORDS GSS
SOURCE Salmomella typhimurium
ORGANISM Salmomella typhimurium
Bacteria; Proteobacteria; gamma subphylum; Enterobacteriaceae;
Salmonella

REFERENCE 1 (bases 1 to 965)

AUTHORS Wong, R.M.-Y., Wong, K.K., Benson, N.R. and McClelland, M.

TITLE Sample sequencing of a Salmonella typhimurium LT2 lambda library:

JOURNAL comparison to the Escherichia coli K12 genome
MEDLINE FEBS Microbiol. Lett. 173 (2), 411-423 (1999)

COMMENT 99243757
Contact: McClelland M
Molecular Biology
Sidney Kimmel Cancer Center
3099 Science Park Road, San Diego, CA 92121, USA
Email: mclelland@fscs.sdsu.edu
Class: Shotgun

FEATURES
source Location/Qualifiers

1..965
/organism="Salmonella typhimurium"
/strain="LT2"
/db_xref="taxon:602"

100

```
/cell_line="RCB-1283 B16 melanoma"  
/note="pooled cell lines : (cell_line=RCB-1751 murmur 100)
```

(cell_line=CRB-2116 JC), (cell_line=RCB-0035 WEHI-3), (cell_line=RCB-0464 Meth-A), (cell_line=RCB-0545 OHR4), (cell_line=RCB-0359 K-1 F1), (cell_line=RCB-1283 B16 melanoma), (cell_type=B cells, cell_line=CR-1702 WEHI 231), (cell_type=TcIdy cells, cell_line=CR-2065 M17C-1), (tissue=Nullipotent stem cell, cell_line=CR-2070 NE), (tissue_type=bladder, cell_line=RCB-0544 MBT-2), (tissue_type=Bone marrow, cell_line=RCB-0544 MBT-2), (cell_line=CR-2028 SR-4967), (tissue_type=colon, cell_line=RCB-0549 C1e-H3), (tissue_type=colon, cell_line=CCL-142 RAG), (tissue_type=kidney, cell_line=CR-1734 SCA-9, clone 15), (strain=AKB/C_{57BL/6}), (cell_type=B cells, cell_line=CR-1669 BC11 clone 13-20-3B3C1), (strain=C3H, tissue_type=brain, cell_line=CR-1443 C3H1).

Query Match	12.18	Score	36.4	DB	10	Length	313
Best Local Similarity	58.28	Pred. No.	0.64				
Matches	64	Conservative	0	Mismatches	46		

QY		Indels	Gaps
47	CCTGCTGAATAATACATCATCATTCAGCTGTCGCGGCTATGCGACTTTAAGTT	0	106
207	CATTACTGATACAGACACCATCTACTGGGACGACCTGGTATACACAGCAGCGCAACAGG	148	

Db 147 TTGACCTACCCACGCTGCTCATGTACTGGCTGCACACACAATG 98

RESULT 4
BE298245

Accession	Definition	Map	EST
601118102F1	NIH_MGC_17 Homo sapiens CDNA clone IMAGE:3027793 5'	LINEAR	EST 20-JUL-2000
mRNA sequence.			
.....			

VERSION
KEYWORDS
SOURCE
human
BE298245.1
GI:9181984
EST.

Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Placentalia

AUTHORS 1 (cases 1 to 690)
TITLE NIH-MGC <http://mgc.ncl.nih.gov/>,
National Institutes of Health

COMMENT
Contact: Robert Strausberg, Ph.D.
Email: c9apbs-rf@mail.nih.gov

CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The T M A C F Consortium

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/INT at: imgc.cba.hawaii.edu

High quality sequence stop: 578.
Location/Qualifiers
1 200

```
/organism="Homo sapiens"  
/db_xref="taxon:9606"  
/gene="BTF2L2"
```

```
/clone_lib="NIH_MGC_17"  
/tissue_type="rhabdomyosarcoma"  
/lab_host="murine"
```

/nucleo=organ: muscle; Vector: pOTB7; Site_1: EcoRI; Site_2: XhoI; cDNA made by oligo-dT priming. Directionally cloned.

following 5' adaptor: GGCACGAG(G). Size-selected for average insert size 1.8kb. Library constructed by Linq Hong in the laboratory of Dr. Y. Zhang.

... (Stratagene) and Superscript II RT (Life Technologies). "

[illegible]

Db 724 GCCTTGATCATTTATCTGCAGAAATTAACATGACTGATPAATATTTCTGTTCATTTTCAG 666

QY 189 GACCCATTTTACATCGTAGCCGATGAGAGCGCCTG 226

664 GTCCTGTAATACATCTGTAGATGTGGAGGCGCTGCTG 627

Db

RESULT 6	829 bp	mRNA	linear	EST 25-SEP-2000
BI763074				
LOCUS	603047876p1	NIH_MGC_116	Homo sapiens	CDNA clone IMAGE:5187906 5'
DEFINITION	mRNA sequence.			
ACCESSION	BI763074			
VERSION	BI763074.1			
KEYWORDS	GI:15754652			
SOURCE	EST.			
ORGANISM	human.			
	Homo sapiens			
	Metazoa:	Chordata:	Cranialata:	Vertebrata:
				Euteleostomi:

FEATURES	Location/Qualifiers
source	1. . 829

BASE COUNT	225 a	219 c	179 g	206 t
ORIGIN				

RESULT 7

AA336439	AA336439	373 bp	mRNA	linear	EST 21-APR-1997
LOCUS					
DEFINITION	EST414118	Endometrial tumor Homo sapiens	CDNA 5'	end,	mRNA sequence
ACCESSION	AA336439				
VERSION	AA336439.1	GI:1988677			
KEYWORDS	EST.				
SOURCE	human.				

BASE COUNT ORIGIN	104 a	91 c	86 g	90 L	2
					2

RESULT 8					
BF727196					
LOCUS					
BF727196	521 bp	mRNA	linear	EST	05-JAN-2001

Email: cgabs-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNLN at:
<http://image.llnl.gov>
 Plate: LLM12033 row: h column: 18
 High quality sequence stop: 622.
 Location/Qualifiers

FEATURES
 source
 1..623
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5405249"
 /clone_lib="NIH_MGC_90"
 /tissue_type="adenocarcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: liver; Vector: PCMV-SPORE6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
 Average insert size 1.7 kb. Library enriched for
 full-length clones and constructed by Life Technologies.
 Note: this is a NIH_MGC Library."
 BASE COUNT 178 a 161 c 138 g 146 t
 ORIGIN

Query Match 11.5%; Score 34.4; DB 13; Length 623;
 Best Local Similarity 53.8%; Pred. No. 3.5;
 Matches 71; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

OY 153 AAGTCCGCTGATTAAGCTTGAAGATTCATTCAGACCCATTTTACATCGTAGCCGA 212
 |||||
 DB 108 AAGGCTGCTGTTGGCTTGAAGATTCATTCATTCACACATTAAGACATGAGC 167
 |||||
 OY 213 TGAGACGCCGCTGATGGTGTCTGCTACCTGACCTGCTCATTTGGAAGCTTACA 272
 |||||
 DB 168 TGTGTAAGACCTTCGATTCCTGCAACACACAGGCTGATGCTCTCTATCA 227
 |||||
 OY 273 TTCTCGCTGATT 284
 |||||
 DB 228 ACCATTCCAATT 239

RESULT 11
 AL529169 728 bp mRNA linear EST 13-FEB-2001
 LOCUS AL529169 LTL_NFL001_NBC4 Homo sapiens CDNA clone CSDD002YE09 5
 DEFINITION
 Prime, mRNA sequence.
 ACCESSION AL529169
 VERSION AL529169.1 GI:12792662
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 728)
 AUTHORS Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
 TITLE Full-length CDNA libraries and normalization
 JOURNAL Unpublished (2001)
 COMMENT Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 Evry cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
 Location/Qualifiers
 1..728
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="CSDD002YE09"
 /clone_lib="LTL_NFL001_NBC4"
 /sex="male"
 /tissue_type="neuroblastoma cells"
 /lab_host="DH10B"
 /note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA"

was primed with a NotI-oligo(dT) primer. Five prime end
 enriched, double-stranded cDNA was digested with Not I and
 cloned into the Not I and Eco RV sites of the pCMVSPORT 6
 vector. Library was normalized. Library was constructed
 by Life Technologies. Contact : Feng Liang Life
 Technologies, a division of Invitrogen 9800 Medical Center
 Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
 8371 Email : fliang@lifestechnology.com
<http://fulllength.invitrogen.com>

BASE COUNT 175 a 201 c 189 g 163 t
 ORIGIN
 Query Match 11.5%; Score 34.4; DB 9; Length 728;
 Best Local Similarity 53.8%; Pred. No. 3.7;
 Matches 71; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

OY 153 AAGTCCGCTGATTAAGCTTGAAGATTCATTCAGACCCATTTTACATCGTAGCCGA 212
 |||||
 DB 398 AAGGCTGCTGTTGGCTTGAAGATTCATTCATTCACACATTAAGACATGAGC 457
 |||||
 OY 213 TGAGACGCCGCTGATGGTGTCTGCTACCTGACCTGCTCATTTGGAAGCTTACA 272
 |||||
 DB 458 TGTGTAAGACCTTCGATTCCTGCAACACACAGGCTGATGCTCTCTATCA 517
 |||||
 OY 273 TTCTCGCTGATT 284
 |||||
 DB 518 ACCATTCCAATT 529

RESULT 12
 BG680800 769 bp mRNA linear EST 01-MAY-2001
 LOCUS BG680800 602628456F1 NCI_CGAP_Skn4 Homo sapiens CDNA clone IMAGE:4753415 5'
 DEFINITION
 mRNA sequence.
 ACCESSION BG680800
 VERSION BG680800.1 GI:13912197
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 769)
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabs-remail.nih.gov
 Tissue Procurement: James Cleaver, M.D.
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNLN at:
<http://image.llnl.gov>
 Plate: LLM10612 row: P column: 24
 High quality sequence stop: 766.
 Location/Qualifiers
 1..769
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4753415"
 /clone_lib="NCI_CGAP_Skn4"
 /tissue_type="squamous cell carcinoma"
 /lab_host="DH10B (T1 phage-resistant)"
 /note="Organ: skin; Vector: PCMV-SPORE6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 1.5kb. Library constructed by Life
 Technologies. Note: this is a NCI_CGAP Library."

FEATURES
 source
 1..769
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4753415"
 /clone_lib="NCI_CGAP_Skn4"
 /tissue_type="squamous cell carcinoma"
 /lab_host="DH10B (T1 phage-resistant)"
 /note="Organ: skin; Vector: PCMV-SPORE6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 1.5kb. Library constructed by Life
 Technologies. Note: this is a NCI_CGAP Library."
 BASE COUNT 181 a 209 c 203 g 176 t
 ORIGIN

Query Match 11.5%; Score 34.4; DB 12; Length 769;
 Best Local Similarity 53.8%; Pred. No. 3.8;

Matches	71	Conservative	0	Mismatches	61	Indels	0	Gaps	0
QY	153	AAAGTCGGCTGTA	AAAGCTTGA	AAAGTTCATTTCCAGACCCATTTTACATCGTAGCCGA	212				
Db	405	AAAGCTCTGGTTTGGCTAGAAAGATCCATTTATATTTACAAACATTAAGACATGAC	464						
QY	213	TGAGAGCGCGCTGATTTGGGTGTTCTGGCTACTGACCTGTCCATTTGTGGAAGCTTTACA	272						
Db	465	TGTTGGAACACCTCTCGATATTTCTCGCAAAACCAACAGGTCTAAGTGCTCTCTCATCA	524						
QY	273	TTCTCGCTGATT	284						
Db	525	ACCATTCCAATT	536						

RESULT	13
LOCUS	B1764142
DEFINITION	B1764142 779 bp mRNA
ACCESSION	F03043378 F1 NIH_MGC_116 Homo sapiens cDNA clone IMAGE:5183587 5'
VERSION	B1764142
KEYWORDS	B1764142.1 GI:15755720 EST.
SOURCE	human.
ORGANISM	Homo sapiens

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 779)		NIH-MGC http://mgc.ncl.nih.gov/ .		
		National Institutes of Health, Mammalian Gene Collection (MGC)		
		Unpublished (1999)		
		Contact: Robert Strausberg, Ph.D.		

FEATURES	Location/Qualifiers
tissue Procurement: Life Technologies, Inc.	
cDNA Library Preparation: Life Technologies, Inc.	
DNA Sequencing by: The I.M.A.G.E. Consortium (LNL)	
clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov	
Plate: LLM11458	row: d column: 20
High quality sequence	step: 767.

BASE COUNT
ORIGIN

183 a 223 c 199 g 174 t

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1 IMAGE:5183587"
/clone_1b="NIH_MGC_116"
/lab_host="DH10B"
/move=Organ: pooled colon, kidney, stomach; Vector:
pcmv-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
source anonymous pool of 3 colons, age 26 yo male, 49 yo
female, 71 yo male colon; 46 yo male kidney, and pool of 2
stomachs, 62 yo male and 70 yo female. Library is
oligo-dr primed and directionally cloned (EcoRV site is
destroyed upon cloning). Average insert size 1.4 kb,
insert size range 1-3 kb. Library is normalized and
enriched for full-length clones and was constructed by C.
Gruber (Invitrogen). Research genetics tracking code
023. Note: this is a NIH_MGC Library."
```

Query Match	Similarity	Score	DB	Length
Best Local	53.8%	53.8%	Pred.	No. 3.8;
Matches	71;	Conservative	0;	Mismatches 61; Indels 0; Caps 0
153	AAAGTCGCTGATTAAGGCTTGA	AAAGTTTCATCTTCAGACCCATTTTACATCGTAGCGGA	212	
1111	1111	1111	1111	1111
424	AAAGCGCTGCTGTTGGCTTGA	GAAGAGTCATTTTATTTACAAACATTAAGACATGAGC	483	
1111	1111	1111	1111	1111
213	TGAGACGGGCTGATTTGGGTCTT	CGGCTACCTGACCTGTCATTTGGTGAAGCTCTTACA	272	
1111	1111	1111	1111	1111

Db	484	TGTTTAAACCCCTCCGTGATATTCCCTGCACCAACAGCTGATGTCCTCTCATCA	543
Oy	273	TTCCTGCTGATT	284
		— — — —	
Db	544	ACCATTCACATT	555

RESULT	14
LOCUS	B1819186
DEFINITION	B1819186 827 bp mRNA linear EST 04-OCT-2001 603036393P1 NIH_MGC_115 Homo sapiens cDNA clone IMAGE:517975 5' , mRNA sequence.
ACCESSION	B1819186
VERSION	B1819186.1 GI:15930736
KEYWORDS	EST.
SOURCE	human.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

koraiyolu; melizoa; choirata; cranata; vertebrata; euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 827)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
email: strausberg@nci.nih.gov

CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
plate: LLAM1443 row: j column: 24
High quality sequence sweep: 824.
location/Qualifiers
1..827

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5177975"
/clone_1fb="NIH_MGC_115"
/lab_host="DH10B"
/notes="Organ: pooled brain, lung, testis; Vector:
pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
source anonymous pool of 6 male brains, age range 23-27; 1
male lung, age 27; and 1 male testis, age 65. Library is
oligo-dT primed and directionally cloned (EcoRV site is
destroyed upon cloning). Average insert size 1.8 kb,
insert size range 1-3 kb. Library is normalized and
enriched for full-length clones and was constructed by C.
Gruber (Invitrogen). Research Genetics tracking code
021. Note: this is a NIH_MGC library."

```

Query Match	11.5%	Score 34.4	DB 13	Length 827
Best Local Similarity	53.8%	Pred. No. 3.9		
Matches	71	Conservative	0	Mismatches 61
				Indels 0
				Gaps
QY	153	AAGTCCGCTGATGAAGGCTTGAAGAAGTTCATTCCAGACCCATTTTTACATCGTAGCCGA	212	
Db	362	AAGGCTGCTGGTTGCTGCTGAAGAGTCCATTATATATTCACCAACATTAAAGACATGAAGC	421	
QY	213	TGAGGACGGCCCTGATGGGTGTTCTGGCTACCTGACCTGTCCATTGTGGAAGGCTTTACA	272	
Db	422	TGTTTGAAGACCCCTCCGATGATTCTGCACAAACCAACAGGCTATATGTCCTCTCTATCA	481	
QY	273	TTCTCGCTGATT	284	
Db	482	ACCATTCCAATT	493	

Fri May 30 09:39:39 2003

us-09-990-099-21.rst

Page 8 -

LOCUS	AU117405	866 bp	mRNA	linear	EST 01-AUG-2002
DEFINITION	AU117405	HEMBA1	Homo sapiens CDNA clone HEMBA1001326 5', mRNA		

DEFINITION	AU117405 HEMBA1 Homo sapiens cdna clone HEMBA1001326 5', mRNA sequence, 5' end
------------	--

ACCESSION	AU117405
VERSION	AU117405.1
	GI:10932371

KEYWORDS	EST.
SOURCE	human

ORGANISM	Homo sapiens
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10	10
11	11
12	12
13	13
14	14
15	15
16	16
17	17
18	18
19	19
20	20
21	21
22	22
23	23
24	24
25	25
26	26
27	27
28	28
29	29
30	30
31	31
32	32
33	33
34	34
35	35
36	36
37	37
38	38
39	39
40	40
41	41
42	42
43	43
44	44
45	45
46	46
47	47
48	48
49	49
50	50
51	51
52	52
53	53
54	54
55	55
56	56
57	57
58	58
59	59
60	60
61	61
62	62
63	63
64	64
65	65
66	66
67	67
68	68
69	69
70	70
71	71
72	72
73	73
74	74
75	75
76	76
77	77
78	78
79	79
80	80
81	81
82	82
83	83
84	84
85	85
86	86
87	87
88	88
89	89
90	90
91	91
92	92
93	93
94	94
95	95
96	96
97	97
98	98
99	99
100	100

ORGANISM	Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.	

REFERENCE
AUTHORS
1 (bases 1 to 866)
Ota, T., Nishikawa, T., Suzuki, Y., Ishii, S., Saito, K., Kawai, Y.,
Nishikawa, Y., Nagai, T., Saito, S., and
Nishikawa, Y.

Yamamoto, J., Wakamatsu, A., Nakamura, Y., Nagai, T., Sugano, S. and Isogai, T.

TITLE HRI human cDNA project
JOURNAL Unpublished (2000)

COMMENT Contact: Takao Isogai

Genomics Laboratory
Helix Research Institute

1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3975

Fax: 81-438-52-3986
Email: genomics@hri.co.jp

HRI human cDNA project; 5' - & 3'-end one pass sequencing; Helix Research Institute; cDNA Library construction; Department of

Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.

FEATURES	Location/Qualifiers
source	1. .866

```

/organism="Homo sapiens"
/db_xref="taxon:9606"

```

```

/clone="HEMBA1001326"
/clone_lib="HEMBA1"

```

```
/tissue_type="whole embryo, mainly head"  
/dev_stage="embryo, 10 weeks"
```

BASE COUNT	219 a	228 c	215 g	201 t	3 others
	/note="Vector: PME18SFL3"				

TABLE NO.	DATE	BY	ORIGIN
1	1950
2	1951
3	1952
4	1953
5	1954
6	1955
7	1956
8	1957
9	1958
10	1959
11	1960
12	1961
13	1962
14	1963
15	1964
16	1965
17	1966
18	1967
19	1968
20	1969
21	1970
22	1971
23	1972
24	1973
25	1974
26	1975
27	1976
28	1977
29	1978
30	1979
31	1980
32	1981
33	1982
34	1983
35	1984
36	1985
37	1986
38	1987
39	1988
40	1989
41	1990
42	1991
43	1992
44	1993
45	1994
46	1995
47	1996
48	1997
49	1998
50	1999
51	2000
52	2001
53	2002
54	2003
55	2004
56	2005
57	2006
58	2007
59	2008
60	2009
61	2010
62	2011
63	2012
64	2013
65	2014
66	2015
67	2016
68	2017
69	2018
70	2019
71	2020
72	2021
73	2022
74	2023
75	2024
76	2025
77	2026
78	2027
79	2028
80	2029
81	2030
82	2031
83	2032
84	2033
85	2034
86	2035
87	2036
88	2037
89	2038
90	2039
91	2040
92	2041
93	2042
94	2043
95	2044
96	2045
97	2046
98	2047
99	2048
100	2049

Query Match	11.5%;	Score 34.4;	DB 9;	Length 866;
Best Local Similarity	53.8%;	Pred. No. 3.9;		

Best Local Similarity	Best Global Similarity	Best Local Alignment	Best Global Alignment
71; Conservative	0; Mismatches	61; Indels	0; Gaps

QY 153 AAAGTCCGCTGATAGCCTTGAAAGTTCATTTCCAGACCCATTTTACATCGTAGCCGA 212

Db 353 AAAGCTGCTGGTTGGCTAGAGAGTCCATTATTTACACACATTAAAGACATGAGC 412

213 TGAGACGGCGCTGATGGGTCTGGCTACTGACCTGTCCATTGTGAAGTCTTACA 272

Db 413 TGTGAAGACCTCTCTGGATATTCTCGCAACCCACAGGTCTATGTGCTCTCTCTATCA 472

Qy 273 TTCTCGCTGATT 284

Db 473 ACCATTCCAATT 484

Search completed: May 5, 2003, 18:41:04
Job time : 1837.49 secs

Search completed: May 5, 2003, 18:41:04
Job time : 1837.49 secs